

OM of: US-09-652-292-2 to: GenEmbl.\* out\_format : pfs

Date: Mar 15, 2002 7:52 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL-frame-p2n.model -drv-xlh
-Q/cgnt2/_USPTO_spool/US09652292/runat_13032002_161725_3207/app_query.fasta_1.606
-DB-GenEmbl -OPMT-fastap -SUFFIX-rge -GAPOP=12.000 -GAPEXT=4.000
-MINNAATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -EGAPOP=6.000
-EGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX-blosum62 -TRANS-human4.0.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTPMT-pfs -NORM-ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09652292_scgnt1.6252
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1
```

Search information block:

Query: US-09-652-292-2

Query length: 541

Database: GenEmbl.\*

Database sequences: 1472140

Database length: -341344837

Search time (sec): 1293.340000

score\_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
gb_pr:HS28H201	+	2765.00	2545.58	1.6e-133	4126	Novel human gene map
gb_pr:AF321240	+	2765.00	2545.50	1.6e-133	4167	AF321240 Homo sapiens Facilita
gb_pr:AF248053	+	2765.00	2545.09	1.7e-133	4396	AF248053 Homo sapiens glucose
gb_pr:HS28H20	+	2286.50	2077.89	1.8e-107	127418	AL031055 Human DNA sequence
gb_hg:AL591064	+	1678.00	1513.69	4.9e-76	200125	AL591064 Mus musculus chromo
gb_hg:AL591805	+	1678.00	1513.17	5.9e-76	243075	AL591805 Mus musculus chromo
gb_pr:AB056798	+	932.00	855.94	2.1e-39	4521	AB056798 Macaca fascicularis b
gb_pr:AL449363	+	727.00	640.15	2.2e-27	141289	AL449363 Human DNA sequence
gb_hg:AC023293	+	724.00	636.25	3.7e-27	163569	AC023293 Homo sapiens clone
gb_ba:BS292954	+	626.50	569.81	1.8e-23	8189	Z9954 B. subtilis yw[A,B,C,D]
gb_ba:BSUB0019	+	626.50	548.36	4.8e-22	212610	Z99122 Bacillus subtilis com
gb_ba:BSUB0021	+	595.00	515.23	2.0e-20	215534	Z99124 Bacillus subtilis com
gb_ba:AB000554	+	592.00	526.35	4.8e-21	36448	AB000554 Lactococcus lactis g
gb_ba:AF045673	+	571.00	521.16	9.4e-21	5957	AF045552 Lactobacillus brevis
gb_pat:AX076672	+	556.00	515.20	2.0e-20	2177	AX076673 Sequence 7 from Paten
gb_pr:HS315644	+	556.00	512.04	3.0e-20	3261	AJ315644 Homo sapiens mRNA for
gb_pat:AX076675	+	547.00	505.82	6.7e-20	2504	AX076675 Sequence 9 from Paten
gb_ro:RNO315643	+	547.00	505.82	6.7e-20	2504	AJ315643 Rattus norvegicus mRN
gb_ba:AB006381	+	541.50	488.47	6.2e-19	12049	AE006381 Lactococcus lactis s
gb_pl:AF280431	+	529.00	489.49	5.5e-19	2422	AF280431 Mesembryanthemum crys
gb_pl:AF280432	+	526.50	488.63	6.1e-19	2013	AF280432 Mesembryanthemum crys
gb_pl:SPTR2	+	520.50	481.91	1.4e-18	2344	X99105 S. pombe ITR2 gene. 7/1
gb_pl:SPAC2068	+	520.50	461.94	1.9e-17	30200	Z99334 S. pombe chromosome I c
gb_pl:AF215852	+	519.00	482.62	1.3e-18	1795	AF215852 Nicotiana tabacum hex
gb_pat:AX172681	+	517.50	480.77	1.7e-18	1905	AX172681 Sequence 171 from Pat
gb_pl:AF215853	+	516.00	480.51	1.7e-18	1650	AF215853 Solanum tuberosum hex
gb_pl:AF280642	+	516.00	480.50	1.7e-18	1653	AX008642 Sequence 3 from Paten
gb_pl:AF280646	+	515.00	477.93	2.4e-18	2041	AF280646 Mesembryanthemum crys
gb_hg:AC093093	+	509.00	440.07	3.1e-16	127914	AC093093 Oryza sativa chromo
gb_hg:AC083945	+	509.00	438.94	3.6e-16	147706	AC083945 Oryza sativa strain
gb_pat:AX008644	+	507.00	471.27	5.7e-18	1864	AX008644 Sequence 5 from Paten
gb_pl:AF215851	+	507.00	471.27	5.7e-18	1864	AF215851 Spinacia oleracea hex
gb_pat:AX120287	+	502.00	468.50	8.1e-18	1473	AX120287 Sequence 203 from Pat
gb_pat:AX120085	+	502.00	425.75	1.9e-15	349980	AX120085 Sequence 1 from Pat
gb_pl:ATC10044	+	501.50	429.71	1.2e-15	198735	AL161544 Arabidopsis thalian
gb_pl:ATFCA6	+	501.50	429.14	1.3e-15	213788	Z97341 Arabidopsis thaliana
gb_ba:ECU00377	+	499.00	449.12	9.7e-17	12354	AE000377 Escherichia coli K12
gb_ba:ECU28377	+	499.00	430.05	1.1e-15	141744	U28377 Escherichia coli K-12
gb_in:LEID1TRA	+	497.50	459.34	2.6e-17	2798	M85072 Leishmania donovani D1
gb_pl:AC006068	+	497.50	432.27	8.4e-16	89338	AC006068 Arabidopsis thaliana

gb\_ba:AE005524 + 496.00 447.13 1.3e-16 11178 AE005524 Escherichia coli  
gb\_ba:AF002563 + 496.00 422.35 3.0e-15 266656 AF002563 Escherichia coli  
gb\_ba:ECOA000368 + 495.00 456.85 3.0e-17 28666 J03732 E.coli arabinoso-pro  
gb\_ba:AE000368 - 495.00 447.24 1.2e-16 9800 AE000368 Escherichia coli K  
gb\_ba:AE005513 - 495.00 446.51 1.4e-16 10755 AE005513 Escherichia coli

seq\_name: gb\_pr:HS28H201

seq\_documentation\_block:

LOCUS HS28H201 4126 bp mRNA PRI 21-NOV-2000  
DEFINITION Novel human gene mapping to chromosome 20, similar to membrane  
transporters.

ACCESSION AL137188

VERSION AL137188.3 GI:11322734

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (24-JAN-2000) E-mail contact: humquery@sanger.ac.uk

On Nov 23, 2000 this sequence was assembled from public domain ESTs and single

pass sequencing reads from expressed DNA templates, aligned to the

genomic DNA sequence from the bacterial clone 28H20 (AL031055). The

EST sequences listed match this sequence with an identity of at

least 95% between the coordinates shown.

Further information can be found at

http://www.sanger.ac.uk/HGP/Chr20/ Sanger Centre name :

dj28H20.C20.1.

FEATURES

source

Location/Qualifiers

1. 4126

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="20"

/map="20q"

1. 112

/number=1

9. 1634

/codon\_start=1

/product="hypothetical protein"

/protein\_id="CAB69822.2"

/db\_xref="GI:11065680"

/translation="MGHSPVPLPCASVSLGGTTFGYELAVISGALLPLQLDPLGLSC

LEQFLVGLSLGALLASLVGGFLIDCVGRKQAILGSLNVLGSLTGLAGSLAWLV

LGRAVGFATLSLSMACCIYVELVGRQGVLSLYEAGITVGSGLSYALNLYALACT

PWGRHFNATAPAVLQSLSLLEFLPAGTDTATHKDLIPLOGSEAPKLGRPRYSF

LDLFRANDNRRTVGLGLVLFPOOLTGPVLCYASTIFSSVFGHSSAVLASVGL

GAVKAAATLTAMGLVDAGRRALLAGCALMALSVSGILVSFAVPMDSGSPCLAVPN

ATGQTGLPGSLQLQDSLLPPIPTNEDQREPIILSTAKTKPHRPSGDPSPAPRLALS

SALGCPPLPARGHALLRWATLCLMVFVSFGFPGVTLVLSIYVPELGRGAFAF

CNSPNAANLPISLFDLIGTIGLSMTFLYGLTAVLGLGFIYLFVPETKGQSLAEI

DQQRKRRTLSFGHRSNRTGIPYSRIEISAS"

13. 1296

/number=2

1297. 1419

/number=3

1338. 1643

/note="matches EST AA313045"

1420. 1555

/number=4

1556. 4126

/number=5

2095. 2137

/note="matches EST AA714011 from clone IMAGE:1240821"

complement(2142. 2172)

/note="matches EST BE747194 from clone IMAGE:3929520"

join(2274. 2544.2548. 2589.2646. 2763)

/note="matches EST AA007343 from clone IMAGE:429263"

complement(2311. 2352)

/note="matches EST AV650272 from clone GICCD12

```

matches EST AV650673 from clone GLCCHC08
matches EST AV650603 from clone GLCCGE10
matches EST AV650406 from clone GLCCGE02
matches EST AV650686 from clone GLCCHD09
join(2417..2711..2814..3001..3029)
/notes="matches EST W02942 from clone IMAGE:291802"
2475..2804
/notes="matches EST W31922 from clone IMAGE:320524"
join(2476..2695..2693..2817)
/notes="matches EST AA489718 from clone IMAGE:823660"
join(2647..2694..2693..3059..3053..3115)
/notes="matches EST AA403072 from clone IMAGE:758347"
join(2739..3054..3050..3143)
/notes="matches EST AA232787 from clone IMAGE:666656"
join(2757..2884..2881..3023)
/notes="matches EST N40351 from clone IMAGE:269966"
complement(join(2794..2876..2894..2935..2935..3045..3046..3138..3135..3249))
/notes="matches EST AA007344 from clone IMAGE:429263"
complement(2796..3249)
/notes="matches EST A1334230 from clone IMAGE:1932055"
2808..2999
/notes="matches EST N44116 from clone IMAGE:272953"
complement(join(2810..3045..3046..3250))
/notes="matches EST A1082631 from clone IMAGE:1660573"
complement(join(2810..2885..2911..2935..3091..3250))
/notes="matches EST N27535 from clone IMAGE:269966"
complement(join(2818..3045..3046..3092..3091..3249))
/notes="matches EST N36110 from clone IMAGE:272953"
complement(join(2905..3045..3046..3138..3135..3249))
/notes="matches EST AA489618 from clone IMAGE:823660"
complement(join(2918..3045..3046..3138..3135..3250))
/notes="matches EST AA233364 from clone IMAGE:666656"
complement(3040..3549)
/notes="matches EST AW298226 from clone IMAGE:2733069"
join(3212..3478..3540..3563)
/notes="matches EST W38959 from clone IMAGE:304871"
complement(3397..3974)
/notes="matches EST AA628914 from clone IMAGE:1032940"
3402..3738
/notes="matches EST C04258 from clone 3NHC3019"
join(3459..3633..3633..3656)
/notes="matches EST AA045035 from clone IMAGE:488780"
3459..3877
/notes="matches EST AAL34031 from clone IMAGE:503848"
join(3542..3710..3754..3948)
/notes="matches EST W39026 from clone IMAGE:305523"
complement(3576..4126)
/notes="matches EST A1042706 from clone IMAGE:1431595"
join(3596..3809..3798..4045)
/notes="matches EST AAL33497 from clone IMAGE:490637"
complement(join(3596..3619..3642..4038))
/notes="matches EST AAL15737 from clone IMAGE:490637"
complement(3597..4126)
/notes="matches EST A1041537 from clone IMAGE:1643803"
complement(3608..3974)
/notes="matches EST A1088144 from clone IMAGE:1683131"
complement(3625..4126)
/notes="matches EST AA404352 from clone IMAGE:758347"
complement(join(3626..3645..3674..3898..3867..4126))
/notes="matches EST AAL33966 from clone IMAGE:503848"
complement(join(3642..3898..3877..4122))
/notes="matches EST AW973035"
complement(join(3685..3898..3867..4126))
/notes="matches EST A1081145 from clone IMAGE:1680293"
complement(join(3689..3898..3867..4126))
/notes="matches EST A1753293 from clone HBMSC_cr08B03"
complement(join(3694..3898..3867..4123))
/notes="matches EST A1097288 from clone IMAGE:1707216"
complement(3727..4126)
/notes="matches EST A1292321 from clone IMAGE:1894742"
complement(join(3737..3898..3867..4126))
/notes="matches EST A1753932 from clone HBMSC_cr16C08"

```

```

misc_feature complement(join(3774..3898..3867..4126))
/notes="matches EST A1277131 from clone IMAGE:1893578"
misc_feature complement(3793..4126)
/notes="matches EST N93207 from clone IMAGE:304871"
misc_feature complement(3795..4124)
/notes="matches EST A1285525 from clone IMAGE:1992441"
misc_feature complement(join(3795..3887..3877..4017..4005..4126))
/notes="matches EST AW103571 from clone IMAGE:2614034"
misc_feature complement(3867..4125)
/notes="matches EST AW166863 from clone IMAGE:2634465"
misc_feature complement(3867..4123)
/notes="matches EST AA532534 from clone IMAGE:986291"
misc_feature complement(3867..4126)
/notes="matches EST A1753418 from clone HBMSC_cr10B04"
misc_feature complement(3945..4126)
/notes="matches EST A1240819 from clone IMAGE:1848499"
BASE COUNT 914 a 1066 c 979 g 1167 t
ORIGIN

alignment_scores:
    Quality: 2765.00      Length: 541
    Ratio: 5.111          Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-652-292-2 x HS28H201

Align seg 1/1 to: HS28H201 from: 1 to: 4126

1 MetGlyHisSerProValLeuProLeuCysAlaSerValSerLeuLeu 17
|||||
9 ATGGGCACTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 58
|||||
17 uGlyGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLeu 34
|||||
59 GGGTGGCCTGACCTTTGGTTATGAACCTGGCAGTCATATCAGGTGCCTGC 108
|||||
34 euProLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnGluPheLeu 50
|||||
109 TGCACATGCAGCTTGACTTTGGGCTAAGCTGCTTGGAGCAGGAGTCCCTG 158
|||||
51 ValGlySerLeuLeuLeuGlyAlaLeuLeuAlaLeuValGlyGlyPh 67
|||||
159 GTGGCAGCCTGCTCTCTGGGGGCTCTCTCTGCTGCTGCTGCTGCTT 208
|||||
67 eLeuIleAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAsnLeu 84
|||||
209 CCTCAPTGACTGCTATGGCAGGAAGCAAGCATCTCGGGAGCAACTTGG 258
|||||
84 aLeuLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaTrp 100
|||||
259 TGCTGCTGCAGGCAGCCTGACCTGGGCTGCTGCTGCTGCTGCTGCTGG 308
|||||
101 LeuValLeuGlyArgAlaValValGlyPheAlaIleSerLeuSerSerMe 117
|||||
309 CTGGTCTCTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
|||||
117 talAcysCysIleTyrValSerGluLeuValGlyProArgGlnArgGly 134
|||||
359 GCGTTCGCTGTATCTACGTCTAGAGCTGCTGGGGCCACGCGACGGGAG 408
|||||
134 aLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeuSer 150
|||||
409 TGCTGGTGTCTCTATGAGGAGCAGCATCACCCTGGGATCTGCTCTCC 458
|||||
151 TyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTyrGlyTyrArgHisMe 167
|||||
459 TATGCCCTCAACTATGCACTGCTGCTACCCCTCGGGATGGAGGCACAT 508
|||||
167 PheGlyTyrPheAlaThrAlaProAlaValLeuGlnSerLeuSerLeuLeu 184
|||||
509 GTTCGGTGGGCGCACTGCACCTGCTGCTGCTGCTGCTGCTGCTGCTCT 558
|||||

```

184 heLeuProhlaGlyThrAspGluThrAlaThrHisLysAspLeuIlePro 200  
 |||||  
 559 TCCTCCCTGCTGCTACAGATGAGCTGCAACACACAAGGACCTCATCCCA 608  
 |||||  
 201 LeuGInGlyGlyGluAlaProLysLeuGlyProGlyArgProArgTyrSe 217  
 |||||  
 609 CTCACGGGAGGTGAGGCCCCCAAGCTGGCCCGGGAGGCCACGGTACTC 658  
 |||||  
 217 rPheLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrV 234  
 |||||  
 659 CTTTCTGGACCTCTTCAGGCGCAGCGATTAACATGCGAGCGCGACACAG 708  
 |||||  
 234 aIGlyLeuGlyLeuValLeuPheGInGInLeuThrGlyGlnProAsnVal 250  
 |||||  
 709 TGGGCGCTGGGCTGCTCTTCCAGCACTAACAGGCGAGGCCAACGCTG 758  
 |||||  
 251 LeuCysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyClySe 267  
 |||||  
 759 CTGTGCTATGCTCCACCATCTTCAGCTCCGTGGTTTCCATGGGGATC 808  
 |||||  
 267 rSerAlaValLeuAlaSerValGlyLeuGlyAlaValLysValAlaAlaT 284  
 |||||  
 809 CTCAGCGCTGCTGGCTCTCTGGGCTTGGCGCAGTGAAGTGGCAGCTA 858  
 |||||  
 284 hrLeuThrAlaMetGlyLeuValAspArgAlaGlyArgArgAlaLeuLeu 300  
 |||||  
 859 CCCTGACCCCATGGGCTGGTGGACGCTGCAGCGCGCAGGCTCTGTGTG 908  
 |||||  
 301 LeuAlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuVa 317  
 |||||  
 909 CTAGCTGGCTGTGCTCTATGCGCTCTCCGTCAGTGGCATAGGCTCTGT 958  
 |||||  
 317 lSerPheAlaValProMetAspSerGlyProSerCysLeuAlaValProA 334  
 |||||  
 959 CAGCTTTGCGCTGCCATGCACTCAGGCCCAAGCTGTCTGGCTGTGCCCA 1008  
 |||||  
 334 snAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAsp 350  
 |||||  
 1009 ATGCCACCGGCGAGACAGCGCTCCCTGGAGACTCTGGCCCTGTGCAAGAC 1058  
 |||||  
 351 SerSerLeuProProIleProArgThrAsnGluAspGlnArgGluProI 367  
 |||||  
 1059 TCCTCTCTACCTCCCATTCACAGGACCAATGAGGACCAAGGAGGACCAAT 1108  
 |||||  
 367 eLeuSerThrAlaLysLysThrLysProHisProArgSerGlyAspProS 384  
 |||||  
 1109 CTTGTCCACTGCTAAGAAACCAAGCCCATCCACATCTGGAGACCCCT 1158  
 |||||  
 384 eAlaProProArgLeuAlaLeuSerSerAlaLeuProGlyProLeu 400  
 |||||  
 1159 CAGCCCTCTCGGCTGGCCCTGAGCTCTGCCCTCCCTGGGCCCTCTGTG 1208  
 |||||  
 401 ProAlaArgGlyHisAlaLeuLeuArgTTrpThrAlaLeuLeuCysLeuMe 417  
 |||||  
 1209 CCGCTCGGGGGGATGACATGCTGCCCTCGACCGCACTGTGCTGCCCTGT 1258  
 |||||  
 417 tValPheValSerAlaPheSerPheGlyPheGlyProValThrTrpLeuV 434  
 |||||  
 1259 GGTCTTTGTGTCAGTGCCTCTCTCTTTGGGTTTGGGCCAGTGACCTGGCTG 1308  
 |||||  
 434 aLeuSerGluIleTyrProValGluIleArgGlyArgAlaPheAlaPhe 450  
 |||||  
 1309 TCCTCAGCAGATACCTCTGTGGAGATACGAGGAAGAGCCCTTCGCTTC 1358  
 |||||  
 451 CysAsnSerPheAsnTrpAlaAlaAsnLeuPheIleSerLeuSerPheLe 467  
 |||||  
 1359 TCAACAGCTTCAACTGGGGGCGCAACCTCTTCATCAGGCTCTCTCTCTCT 1408  
 |||||  
 467 uAspLeuIleGlyThrIleGlyLeuSerTrpThrPheLeuLeuTyrGlyL 484  
 |||||  
 1409 CGATCTCATTTGGCACCATCGGCTTGTCTGGACCTTCTCTCTCTCTCTCT 1458  
 |||||

484 euThrAlaValLeuGlyLeuGlyPheIleTyrLeuPheValProGluThr 500  
 |||||  
 1459 TGACCCCTCTCTCGGCTGGCTTCATCTATTTATTTCTCTCTCTCTCTCT 1508  
 |||||  
 501 LysGlyGlnSerLeuAlaGluIleAspGlnGlnPheGlnLysArgArgPh 517  
 |||||  
 1509 AAGGCCAGTCTGTTGGCAGATAGACACAGCTTCCAGAAAGACAGGTT 1558  
 |||||  
 517 eThrLeuSerPheGlyHisArgGlnAsnSerThrGlyIleProTyrSerA 534  
 |||||  
 1559 CACCCTGAGCTTGGCCACAGGCGAAGCTCCACTGGCATCCCGTACAGCC 1608  
 |||||

534 rgIleGluIleSerAlaAlaSer 541

1609 GCATCAGATCTCTCGGCTCC 1631

seq\_name: gb\_pr:AF321240

seq\_documentation\_block: 4167 bp mRNA PRI 25-MAR-2001

LOCUS AF321240  
 DEFINITION Homo sapiens facilitative glucose transporter GLUT10 (SLC2A10)  
 mRNA, complete cds.

ACCESSION AF321240

VERSION AF321240.1 GI:13445574

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

1. 4167

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="20"

/map="20q13.1"

1. 4167

/gene="SLC2A10"

53. 1678

/gene="SLC2A10"

/note="member of the facilitative glucose transporter family"

/codon\_start=1

/product="facilitative glucose transporter GLUT10"

/protein\_id="AAK26294.1"

/db\_xref="gi:13445574"

/translation="MGHSPVPLCASVSLGLTGFYELAVISGALLPLQLDFGLSC

LEQFLVGLSLGLLALSLVGGFLIDCYGRKQAILGSLNLLAGSLTGLAGSLAWLV

LGRAVGFPAISLSMACCIVYSELVGPQRGVLSLYEAGITVGTLLSYALNYALACT

PWGRHMFGNATAPVQLSLSLFLPAGTDETATKHLLIPLOGGEAPKLGPRPRYSF

LDLFRANDNHRRTTVGLVLPFOOLTGPQNVLCYASTIFSSVGFHGGSSAVLASVGL

GAVKAATLTAMGLVDVDRAGRALLAGCMLASVSGIGLVSPFVPMDSGSPCLAVPN

ATGOTGLPGDSGLLQDSLPPIPTNEDREPIILSTAKTKLPHRPSGDPSPAPRLALS

SALPGPLPARGHALLRWTLCLMVFVSASFSGFPGVTWLVLEIYVETRGRAFAP

CNSFNWAANLPLSLDFLDLIGTISWTFLLYGLTAVLGLGFTILFVLPETGQSLAEI

DQFQKRFTLSPGHRNSTGIPISREISAAS"

BASE COUNT 916 a 1090 c 991 g 1170 t

ORIGIN

alignment\_scores: 2765.00 Length: 541

Quality: 2765.00

Ratio: 5.111 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-652-292-2 x AF321240  
 Align seg 1/1 to: AF321240 from: 1 to: 4167

```

1 MetGlyHisSerProProValLeuProLeuCysAlaSerValSerLeuLeu 17
|||||
53 ATGGCCCACTCCCAACCTGCTCCTGCTTGTGTGCTGCTGCTGCTGCT 102
|||||
17 uClyGlyLeuThrPheGlyTyrGlnLeuAlaValIleSerGlyAlaLeuL 34
|||||
103 GGGTGGGCTGACCTTTGGTTATGAACATGGCAGTATATATATATATAT 152
|||||
34 euProLeuGlnLeuAspPheGlyLeuSerCysLeuGlnGluPheLeu 50
|||||
153 TGCCACTGCAGCTTGACTTTGGGCTAGCTGCTTGGAGCAGGAGTTCCTG 202
|||||
51 ValGlySerLeuLeuGlyAlaLeuLeuAlaSerLeuValGlyGlyPh 67
|||||
203 GTGGCAGCCTGCTCCCTGGGGCTCTCCTGCCTCCCTGCTGCTGCT 252
|||||
67 eLeuIleAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAsnLeuV 84
|||||
253 CCTCATTTGACTGCTATGGCAGGAAGCAAGCAATCCTCGGGAGCAACTGG 302
|||||
84 aLeuLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaTri 100
|||||
303 TGCTGCTGGCAGGAGCAGCTGACCTGGGCTGCTGCTGCTGCTGCTGCT 352
|||||
101 LeuValLeuGlyArgAlaValValGlyPheAlaIleSerLeuSerMe 117
|||||
353 CTGGTCTCGGGCCGCGTGTGGTGGCTTCGCCATTTTCCCTCTCTCCAT 402
|||||
117 tAlaCysCysIleTyrValSerGluLeuValGlyProArgGlnArgGlyV 134
|||||
403 GGCTGTGTATCTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 452
|||||
134 aLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeuSer 150
|||||
453 TGCTGGTGTCCCTCTATGAGCAGGCATCACCGTGGGCATCTCTGCTCC 502
|||||
151 TyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTrpGlyTrpArgHisMe 167
|||||
503 TATGCCCTCAACTATGCACTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 552
|||||
167 tPheGlyTrpAlaThrAlaProAlaValLeuGlnSerLeuSerLeuLeuP 184
|||||
553 GTTCGGCTGGCCACTGCACCTGCTGTCTGCAATCCCTCAGCCTCTCTCT 602
|||||
184 heLeuProAlaGlyThrAspGluThrAlaThrHisLysAspLeuIlePro 200
|||||
603 TCCTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 652
|||||
201 LeuGlnGlyGlyGluAlaProLysLeuGlyProGlyArgProArgTyrSe 217
|||||
653 CTCAGGAGGTGAGGCCCCCAAGCTGGGGCCGGGAGGCCGACTC 702
|||||
217 rPheLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrThrV 234
|||||
703 CTTTCTGGACCTCTTACGGGCACCGCATACATGCGAGGCCGACACAG 752
|||||
234 aGlyLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnVal 250
|||||
753 TGGGCTGGGGCTGGTGGCTCTTCACCACTTAACAGGCGCAGCCCAACGTG 802
|||||
251 LeuCysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyGlySe 267
|||||
803 CTGTGTATGCTTCCACCATCTTCAGCTCGTGTGTGTGTGTGTGTGTGTGT 852
|||||
267 rSerAlaValLeuAlaSerValGlyLeuGlyAlaValLysValAlaAlat 284

```

seq\_name: gb\_pr:AF248053

seq\_documentation\_block:

LOCUS AF248053 4396 bp mRNA

DEFINITION Homo sapiens glucose transporter (GLUT10) mRNA, complete cds.

PRI

12-APR-2001

ACCESSION AF248053

```

853 CTCAGCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 902
|||||
284 hrLeuThrAlaMetGlyLeuValAspArgAlaGlyArgArgAlaLeuLeu 300
|||||
903 CCCTGACCGCCATGGGCTGGTGGACCGTGCAGGCCGACGAGGCTCTGTGTG 952
|||||
301 LeuAlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuVa 317
|||||
953 CTAGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1002
|||||
317 lSerPheAlaValProMetAspSerGlyProSerCysLeuAlaValProA 334
|||||
1003 CAGCTTTGGCGTGGCCATGGACTCAGGCCCAAGCTGTCTGCTGTGCCCA 1052
|||||
334 snAlaIleThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAsp 350
|||||
1053 ATGCCACCGGGCAGACAGAGGCTCCCTGGAGACTCTGGCCTGCTGCAGGAC 1102
|||||
351 SerSerLeuProProIleProArgThrAsnGluAspGlnArgGluProIl 367
|||||
1103 TCCTCTCTACCTCCCATTTCCAGGACCAATGAGGACCAAGAGGAGCCCAAT 1152
|||||
367 eLeuSerThrAlaLysLysThrLysProHisProArgSerGlyAspProS 384
|||||
1153 CTTGTCCACTGCTAAGAAACCAAGCCCATCCAGATCTGGAGACCCCT 1202
|||||
384 erAlaProProArgLeuAlaLeuSerSerAlaLeuProGlyProLeu 400
|||||
1203 CAGCCCTCTCTCGGCTGGGCTGAGCTCTGCCCTGCCCTGGGCCCCCTCTG 1252
|||||
401 ProAlaArgGlyHisAlaLeuLeuArgTyrThrAlaLeuLeuCysLeuMe 417
|||||
1253 CCCGCTCGGGGATGACACTGCTGGCTGGACCGCACTGCTGTGCTGCTGAT 1302
|||||
417 tValPheValSerAlaPheSerPheGlyPheGlyProValThrTrpLeuV 434
|||||
1303 GGTCTTTGTTCAGTGCCTTCTCTTTGGTGTGGGCCAGTGCCTGGGCTTG 1352
|||||
434 alLeuSerGluIleTyrProValGluIleArgGlyArgAlaPheAlaPhe 450
|||||
1353 TCCTCAGCGAGATCTACCTGTGGAGATACGAGGAGAGAGCCTTCGCGCTC 1402
|||||
451 CysAsnSerPheAsnTrpAlaAlaAsnLeuPheIleSerLeuSerPheLe 467
|||||
1403 TGCAACAGCTTCAACTGGCGGCCCAACCTCTTCATCAGCCTCTCTCTCT 1452
|||||
467 uAspLeuIleGlyThrIleGlyLeuSerTrpThrPheLeuLeuTyrGlyL 484
|||||
1453 CGATCTCATTGGCACCATCGGCTGTCTGGACCTTCTCTGCTTACGGAC 1502
|||||
484 euThrAlaValLeuGlyLeuGlyPheIleTyrLeuPheValProGluThr 500
|||||
1503 TGACCGCTGCTCTCGGCTGGGCTTCTATCTATTTATTTGTTCTCTGAACA 1552
|||||
501 LysGlyGlnSerLeuAlaGluIleAspGlnGlnPheGlnLysArgArgPh 517
|||||
1553 AAAGGCCAGTCTGTGGCAGATAGACCAGCAGTTCACAGAAAGAGACGGTT 1602
|||||
517 eThrLeuSerPheGlyHisArgGlnAsnSerThrGlyIleProTyrSera 534
|||||
1603 CACCCTGAGCTTTGGCCACAGGACAGACTCCACTGGCAGCTCCCTACAGCC 1652
|||||
534 rgIleGluIleSerAlaAlaSer 541
|||||
1653 GCATCGAGATCTCTGGGCTCC 1675

```

```

VERSION      AF248053.1  GI:13603726
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 4396)
AUTHORS      Foisey,S.C., Mihic,S.J., Craddock,A.L., Mychaleckyj,J.C.,
             Dawson,P.A. and Bowden,D.W.
TITLE        GLUT10: A novel glucose transporter in the type 2 diabetes' linked
             region of chromosome 20q12-13.1
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 4396)
AUTHORS      Dawson,P.A., Mihic,S.J., Craddock,A.L., Mychaleckyj,J.C.,
             Foisey,S.C. and Bowden,D.W.
TITLE        Direct Submission
JOURNAL      Submitted (22-MAR-2000) Molecular Genetics, Wake Forest University,
             Medical Center Boulevard, Winston-Salem, NC 27106, USA
FEATURES     Location/Qualifiers
             1..4396
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="20"
             /map="20q12-q13.1"
             1..4396
             /gene="GLUT10"
             251..1876
             /gene="GLUT10"
             /codon_start=1
             /product="glucose transporter"
             /protein_id="AAK31911.1"
             /db_xref="GI:1360372"
             /translation="MGHSPVPPLICASVSLGGTLFGYELAVISGALLPLQLDLFGLSC
             LEQFLVGLSLGLLGLALLSLVGGFLIDCYGRKOAILGSLNVLGLSLTLGLAGSLAWLV
             LGRVAVGFATISLSMACCIYVSELVGPQRGVLSLYEAGITVGLISLYALNYALAGT
             PWGRHFGWATAPAVLOSLSLFLPACTDETATPHKDLIPLOGGEAPKLGPGRPYSP
             LDLEFRADNRGRTYVGLGLVLFQOLGQPNVLCYASTIFSVEFHGSSAVLASVGL
             GAVVATATAMGLVDKAGRRALLLGLAGCALMALSVSGIGLVSAFVPMDSGSPSLAVPN
             ATGQPLGDSGLLQDSLPILPTNEDQREPIILSTAKKTKPHRPSGDSAPPRIALS
             SCLPPLPARGHALLRWLTALCLMVFVSFSGFVPTWLVLSIYVPEIRGRAFAF
             CNSFNAAALPISLSLDLIGTIGLSWTFLLYGLTAVLGLGFVLFVPETKGQSLAEI
             DOQPKRRFTLSFGHRNSTGIPYSRIEISAS"
BASE COUNT   939 a 1139 c 1102 g 1216 t
ORIGIN

alignment_scores:
  Quality: 2765.00      Length: 541
  Ratio: 5.111          Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-652-292-2 x AF248053
  Align seg 1/1 to: AF248053 from: 1 to: 4396
  1 MetGlyHisSerProProValLeuProLeuCysAlaSerValSerLeuLeu 17
  251 ATGGGCCACTCCACCTGCTCCCTGCTTGTGGCTCTGCTGCTTGTGCT 300
  17 uGlyGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLeuL 34
  301 GGCTGGCCCTGACCTTTGGTTATGAACCTGGCAGTCATATCAGGTGCCCTG 350
  34 euProLeuGlnLeuAspPheGlyLeuSerCysLeuGlnGluPheLeu 50
  351 TGGCACTGACGCTTGAATTTGGGCTAAGCTGCTTGGAGCAGGAGTTCCTG 400
  51 ValGlySerLeuLeuLeuGlyAlaLeuLeuAlaSerLeuValGlyGly 67
  401 GTGGGAGGCTGCTCCTGGGGGCTCTCCTCGCTCCCTGCTGCTGCTGCT 450
  67 eLeuIleAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAsnLeu 84

```

```

451 CCTCATTTGACTGCTATGGCAGGAAGCAAGCATCTCTCGGAGCAACTTGG 500
84 aLeuLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaTrp 100
501 TGTGCTGGCAGCAGCAGCTGACCTGGGCTGGCTGCTGCTGCTGCTGCTG 550
101 LeuValLeuGlyArgAlaValValGlyPheAlaIleSerLeuSerSerMe 117
551 CTGCTCTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
117 tAlaCysCysIleTyrValSerGluLeuValGlyProArgGlnArgGly 134
601 GGCTTGTGCTATCTACGCTGTCAGAGCTGGTGGGCGCACGGCAGCGGGAG 650
134 aLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeuSer 150
651 TGTGCTGCTGCTCTATGAGCAGCAGCATCACCTGGGCTGCTGCTGCTGCT 700
151 TyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTrpGlyTrpArgHisMe 167
701 TATGCCCTCAACTATGCACTGGCTGGTACGCCCTGGGATGGAGGCACAT 750
167 tPheGlyTrpAlaThrAlaProAlaValLeuGlnSerLeuSerLeuLeuP 184
751 GTTGGGCTGGGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 800
184 heLeuProAlaGlyThrAspGluThrAlaThrHisLysAspLeuIlePro 200
801 TCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850
201 LeuGlnGlyGlyGluAlaProLysLeuGlyProGlyArgProArgTyrSe 217
851 CTCAGGAGGTGAGGCGCCCAAGCTGGCGCGGGAGGCGGACGCTACTC 900
217 rPheLeuAspLeuPheArgAlaAlaArgAspAsnMetArgGlyArgThr 234
901 CTTTCTGGAGCTCTTTCAGGCGACCGCATACATCGAGCGCGGACACAG 950
234 aGlyLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnVal 250
951 TGGGCTGGGCTGGTGGTCTCTTCCAGCACTAACAGGCGCAGCCCACTG 1000
251 LeuCysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyGlySe 267
1001 CTGTGCTATGCTCCACCATCTTCAGCTGCTGCTTTCATGGGGATC 1050
267 rSerAlaValLeuAlaSerValGlyLeuGlyAlaValIleValAlaAla 284
1051 CTCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1100
284 hrLeuThrAlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeu 300
1101 CCCTGACCGCATGGGCTGGTGGACCGCTGAGCGCGCAGGGCTCTGTTG 1150
301 LeuAlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuVa 317
1151 CTAGCTGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
317 lSerPheAlaValProMetAspSerGlyProSerCysLeuAlaValProA 334
1201 CAGCTTTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1250
334 snAlaThrGlyGlnThrGlyLeuProGlyLeuAspSerGlyLeuGlnAsp 350
1251 ATGCCACCGGCGACAGAGGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCT 1300
351 SerSerLeuProProIleProArgThrAsnGluAspGlnArgGluProIl 367
1301 TCCTCTCTACCTCCCATTTCCAGGACCAATCAGGACCAAGAGGAGCAAT 1350
367 eLeuSerThrAlaLysLysThrLysProHisProArgSerGlyAspProS 384

```

1351 CTGTGCTCCTCTAGAAAACCAAGCCCCATCCAGATCTGGAGACCCCT 1400

384 erAlaProArgLeuAlaLeuSerSerAlaLeuProGlyProProLeu 400

1401 CAGCCCTCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1450

401 ProAlaArgGlyHisAlaLeuLeuLeuArgTrpThrAlaLeuLeuCysLeu 417

1451 CCCGCTCGGGGGCATGCACTCTCGCTGCTGCTGCTGCTGCTGCTGCTG 1500

417 tValPheValSerAlaPheSerPheGlyPheGlyProValThrTrpLeu 434

1501 GGTCTTGTCTAGTCCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1550

434 alLeuSerGluIleTyProValGluIleArgGlyArgAlaPheAlaPhe 450

1551 TCCTCAGGAGATCTACCTCTGTGAGATACGAGGAGAGCTTCGCTTC 1600

451 CysAsnSerPheAsnTrpAlaAlaAsnLeuPheLeuSerLeuSerPheLe 467

1601 TGCACAGCTTCAACTGGGGGCGCAACCTCTTCATCAGCCCTCTCTCT 1650

467 uAspLeuIleGlyThrIleGlyLeuSerTrpThrPheLeuLeuTyGlyL 484

1651 CGATCTCATTTGGCAACATCGCTTGTCTCTGACCTCTCTCTCTACG 1700

484 euThrAlaValLeuGlyLeuGlyPheIleTyLeuPheValProGluThr 500

1701 TGACCGCTGCTCTGGCGCTGGCTTCATCTATTATTGTTCTTCAACA 1750

501 LysGlyClnSerLeuAlaGluIleAspGlnGlnPheGlnLysArgArgPh 517

1751 AAAGCCAGTCGTTGGCAGAGATAGACGACGAGTTCACAGAGACGGTT 1800

517 eThrLeuSerPheGlyHisArgGlnAsnSerThrGlyIleProTySerA 534

1801 CACCTCAGCTTGGCCACAGGAGAACTCCACTGGCATCCCGTACAGCC 1850

534 rgIleGluLeuSerAlaAlaSer 541

1851 GCATCAGAGATCTCTGGCGGCTCC 1873

seq\_name: gb\_pr:HS28H20

seq\_documentation\_block: 19-APR-2001 PRI  
 LOCUS HS28H20 127418 bp DNA  
 DEFINITION Human DNA sequence from clone RPL-28H20 on chromosome 20q13.1  
 Contains the SLC2A10 gene encoding a solute carrier family 2  
 (facilitated glucose transporter) member 10, the 5' end of a novel  
 gene, ESTs, STSS, GSSs and three CpG islands, complete sequence.

ACCESSION AL031055.1 GI:4375937  
 VERSION AL031055.1  
 KEYWORDS HTG; CpG Island; SLC2A10.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 127418)  
 Ramsay, H.  
 Direct Submission

COMMENT  
 Submitted (18-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerquests@sanger.ac.uk  
 On Mar 7, 1999 this sequence version replaced gi:4056528.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information

on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 20, constructed by the Sanger Centre Chromosome 20  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr20  
 This sequence is the entire insert of clone RPL-28H20 This sequence  
 was finished as follows unless otherwise noted: all regions were  
 either double-stranded or sequenced with an alternate chemistry or  
 covered by high quality data (i.e., phred quality >= 30); an  
 attempt was made to resolve all sequencing problems, such as  
 compressions and repeats; all regions were covered by at least one  
 plasmid subclone or more than one M13 subclone; and the assembly  
 was confirmed by restriction digest. RPL-28H20 is from the library  
 RPL-1 constructed by the group of Pieter de Jong. For further  
 details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pCYPAC2.

# FEATURES

source  
 1..127418  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /map="q13.1"  
 /clone="RPL-28H20"  
 /clone\_lib="RPL-1"  
 complement(join(1..562,2463..3110))  
 /gene="dJ101A2.2"  
 /note="match: CDNAS: Em:AK013049 Em:AK011574  
 match: ESTs: Em:AA233726 Em:T87841 Em:AA756018 Em:AA113277  
 Em:AA985248 Em:AA682750 Em:AA61487 Em:AI028588  
 Em:AA659710 Em:AT1149891 Em:AA579641 Em:AA516077  
 Em:AA460132 Em:AA483490 Em:AA046575"  
 /evidence="not\_experimental"  
 /product="dJ28H20.2 (novel protein)"  
 complement(1..3110)  
 /gene="dJ101A2.2"  
 18..494  
 /note="match: STS: Em:HS28H20T"  
 complement(join(84..562,2463..2745))  
 /gene="dJ101A2.2"  
 /note="supported by GENESH and GENSCAN  
 continues in Em:AL133520 as dJ101A2.2"  
 /codon\_start=1  
 /evidence="not\_experimental"  
 /product="dJ28H20.2 (novel protein)"  
 /protein\_id="CAC00561.1"  
 /db\_xref="GI:9588402"  
 /db\_xref="SPTREMBL:O9NOE6"  
 /translation="MAAARATTPADGEEPAEAALAAARSSRPLSLGLVKGQAE  
 ARVFRGFGRAAVTKHRPKGYRHPALERLGRRTVQEARLLRCRAGISAPVVE  
 FVDYASNCLYMEIEGYSVTROYIOSTMETKTPQGLSLAKTIQVLRMHDEDLIH  
 GDLTTSNMLLKPPLEQLNVLIDFGLSFISALPDKGVGLYVLEKAFLLSTHPNTETVE  
 EAFKASYSLLTSKKARPVLKLDKDELGRGRKSMVG"  
 867..1145  
 /note="AluX repeat: matches 1..298 of consensus"  
 1251..1353  
 /note="MIR repeat: matches 50..151 of consensus"  
 1672..1797  
 /note="MIR repeat: matches 8..135 of consensus"  
 1855..2079  
 /note="MIR repeat: matches 33..262 of consensus"  
 2155..2248  
 /note="MIR repeat: matches 33..135 of consensus"  
 2384..3150  
 /note="CpG island"  
 /evidence="not\_experimental"  
 2812..2871  
 /note="3 copies 20 mer 85% conserved"  
 complement(2876..3110)  
 /gene="dJ101A2.2"  
 /note="match: STS: Em:GL15621"  
 3865..3962

repeat\_region

```

repeat_region /note="MER21B repeat: matches 299. .391 of consensus"
3963. 5655
repeat_region /note="L1M8 repeat: matches 4532. .6270 of consensus"
5669. 5895
repeat_region /note="L1ME repeat: matches 5501. .5739 of consensus"
6332. 6548
repeat_region /note="L2 repeat: matches 2524. .2746 of consensus"
6626. 6915
repeat_region /note="AluX repeat: matches 3. .290 of consensus"
6918. 6980
repeat_region /note="MER94 repeat: matches 1. .62 of consensus"
7023. 7322
repeat_region /note="AluJ repeat: matches 1. .312 of consensus"
7323. 7417
repeat_region /note="L2 repeat: matches 2389. .2513 of consensus"
7457. 7623
repeat_region /note="MIR repeat: matches 93. .262 of consensus"
7864. 7934
repeat_region /note="MLT1 repeat: matches 311. .382 of consensus"
8046. 8240
repeat_region /note="MLT1J repeat: matches 9. .211 of consensus"
8347. 8647
repeat_region /note="AluS repeat: matches 1. .300 of consensus"
8685. 8793
repeat_region /note="MLT1J repeat: matches 271. .366 of consensus"
9095. 9221
repeat_region /note="MIR repeat: matches 102. .242 of consensus"
9647. 9776
repeat_region /note="L2 repeat: matches 2170. .2303 of consensus"
9793. 10028
repeat_region /note="AluJ repeat: matches 74. .306 of consensus"
10038. 10206
repeat_region /note="LTR16C repeat: matches 126. .316 of consensus"
10631. 10690
repeat_region /note="MLT1D repeat: matches 445. .505 of consensus"
10691. 10995
repeat_region /note="AluJ repeat: matches 1. .295 of consensus"
10996. 11200
repeat_region /note="MLT1D repeat: matches 248. .445 of consensus"
11232. 11634
repeat_region /note="MSTB repeat: matches 2. .409 of consensus"
11670. 11761
repeat_region /note="MLT1D repeat: matches 128. .192 of consensus"
11762. 12062
repeat_region /note="AluX repeat: matches 1. .293 of consensus"
12063. 12180
repeat_region /note="MLT1D repeat: matches 1. .128 of consensus"
complement(12181. 12344)
misc_feature /note="match: GSS: Em:AQ190762"
13011. 13129
repeat_region /note="MLT1J repeat: matches 252. .368 of consensus"
13210. 13392
repeat_region /note="MER20 repeat: matches 1. .215 of consensus"
13397. 13442
repeat_region /note="MIR repeat: matches 104. .146 of consensus"
13660. 13783
repeat_region /note="LTR33 repeat: matches 330. .452 of consensus"
13894. 14254
repeat_region /note="L2 repeat: matches 2255. .2619 of consensus"
14305. 14435
repeat_region /note="L2 repeat: matches 2617. .2748 of consensus"
14446. 14525
repeat_region /note="4 copies 20 mer 76% conserved"
15476. 15752
repeat_region /note="AluJ repeat: matches 2. .278 of consensus"
15774. 15952
repeat_region /note="MER58A repeat: matches 42. .220 of consensus"
16190. 16304
repeat_region /note="L2 repeat: matches 2657. .2744 of consensus"
16305. 16599
repeat_region /note="AluJ repeat: matches 1. .296 of consensus"
16600. 16635
repeat_region /note="L2 repeat: matches 2617. .2657 of consensus"

```

```

repeat_region 16641. 16803
/note="L1M10 repeat: matches 6162. .6284 of consensus"
repeat_region 16804. 17190
/note="MER57B repeat: matches 1. .403 of consensus"
repeat_region 17198. 17297
/note="5 copies 20 mer 71% conserved"
repeat_region 17204. 17297
/note="47 copies 2 mer ta 70% conserved"
repeat_region 17559. 17868
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 17939. 18240
/note="LTR26 repeat: matches 246. .553 of consensus"
misc_feature complement(18333. 18409)
/note="match: SFS: Em:G42885"
repeat_region 18451. 18615
/note="FAM repeat: matches -126. .41 of consensus"
repeat_region 19258. 19336
/note="MER57-internal repeat: matches 7168. .7244 of consensus"
repeat_region 19343. 19485
/note="MER53 repeat: matches 2. .180 of consensus"
repeat_region 19721. 20190
/note="LTR8 repeat: matches 251. .691 of consensus"
repeat_region 20238. 20493
/note="LTR8 repeat: matches 1. .256 of consensus"
repeat_region 20704. 20978

alignment_scores:
  Quality: 2286.50      Length: 651
  Ratio: 4.855         Gaps: 2
  Percent Similarity: 72.350  Percent Identity: 72.197

alignment_block:
US-09-6527292-2 x HS28H20 ..

Align seg 1/1 to: HS28H20 from: 1 to: 127418

1 MetGlyHisSerProProValLeuProLeuCysAlaSerValSerLeuLe 17
:::|||||
38368 TTAGGCCACCTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 38417

17 uGlyGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLeu 34
|||||
38418 GGGTGGCTGACCTTTGGTATGAACGGCAGTCATATCAGTGCCTGC 38467

34 euProLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnClupheLeu 50
|||||
38468 TGGCAGCTGACCTTTGGGCTTACCTGCTTGGAGCAGGAGTTCCTG 38517

51 ValGlySerLeuLeuGlyAlaValAlaLeuAlaSerLeuValGlyGlyph 67
|||||
38518 GTGGGAGCCTGCTCTCTGGGGCTCTCTCTGCTCTCTCTGCTGCTT 38567

67 eLeuIleAspCysTyrGlyArgLysGlnAlaIleLeuGlySerAnLeuV 84
|||||
38568 CCTCATTTGACTGCTATGCGAGGAAGCAAGCATCTCGGGAGCACTGG 38617

84 alLeuLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAlaTrp 100
|||||
38618 TGCTGCTGGGAGGAGCCTGACCTGGGCTGGCTGGCTGGCTGGCTGG 38667

101 LeuValLeuGlyArgAlaValAlaValGlyPheAlaIleSerLeuSerMe 117
|||||
38668 CTGCTCTGGGCGCGCTGCTGGTGGCTTGGCTTGGCTTCCCTCTCCAT 38717

117 tAlaCysCysIleTyrValSerGluLeuValGlyProArgGlnArgGlyV 134
|||||
38718 GGCTTGCTGTATCTACGTGTGTCAGAGCTGGTGGGCGCAGCGCGGGAG 38767

134 alLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeuSer 150
|||||
38768 TGGTGGTGTCCCTCTATGAGGAGGAGCATCACCGTGGGATCCTGCTCTCC 38817

```



151 TrrAlaLeuAsnTrrAlaLeuAlaGlyThrProTrrPglyTrrPArgHisMe 167  
|||||  
38818 TATGCCCTCACTATGACCTGGTGTACCCCTGGGATGGAGCACAT 38867  
167 tPheGlyTrrAlaThrAlaProAlaValLeuGlnSerLeuSerLeuLeuP 184  
|||||  
38868 GTTCGGCTGGCCACTGCACCTGCTGCCATCCCTCAGCCCTCCTCT 38917  
184 heLeuProAlaGlyThrAspGluThrAlaThrHisLysAspLeuIlePro 200  
|||||  
38918 TCCTCCCTGCTGTACAGTACGACTGCAACACACAGGACCTCATCCCA 38967  
201 LeuGlnGlyGluAlaProLysLeuGlyProGlyArgProArgTyrSe 217  
|||||  
38968 CTCAGGAGGTGAGGCCCCCAAGCTGGGCCCGGGAGGCGACGGTACTC 39017  
217 rPheLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrV 234  
|||||  
39018 CTTCTGGACCTCTTACGGGACGCGATAACATGCGAGGCGCGACACAG 39067  
234 alGlyLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnVal 250  
|||||  
39068 TGGCCCTGGGGCTGGTCTCTTCAGCACTAACAGGGCAGCCCAACGTG 39117  
251 LeuCystTrrAlaSerThrIlePheSerSerValGlyPheHisGlyGlySe 267  
|||||  
39118 CTGTGTATGCCCTCCACCATCTTCAGCTCGGTTGGTTCCATGGGGATC 39167  
267 rSerAlaValLeuAlaSerValGlyLeuGlyAlaValLysValAlaAlaT 284  
|||||  
39168 CTCAGCGGTGCTGGCTCTCTGGGGCTTGGCGAGTGAAGGTGGCAGCTA 39217  
284 hrLeuThrAlaMetGlyLeuValAspArgAlaGlyArgArgAlaLeuLeu 300  
|||||  
39218 CCCTGACCGCATGGGGCTGTGGACCGTGCAGGCCGCGAGGCTCTGTG 39267  
301 LeuAlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuVa 317  
|||||  
39268 CTAGCTGGCTGTGCCCTCATGGCCCTGTCCGTGAGTGGCAGTGGCTCGT 39317  
317 lSerPheAlaValProMetAspSerGlyProSerCysLeuAlaValProA 334  
|||||  
39318 CAGCTTGGCTGCCATGACCTAGCCCAAGCTGCTGGCTGGTGGCCCA 39367  
334 snAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAsp 350  
|||||  
39368 ATGCCACGGGACAGAGGCTCCCTGGAGACTCTGGCTGCTGCAGGAC 39417  
351 SerSerLeuProProIleProArgThrAsnGluAspGluProII 367  
|||||  
39418 TCCTCTCTACCTCCCATTTCCAGGACCAATGAGGACCAAGGAGGCAAT 39467  
367 eLeuSerThrAlaLysLysThrLysProHisProArgSerGlyAspProS 384  
|||||  
39468 CTTGTCCACTGCTAGAAAACCAAGCCCATCCAGATCTGGAGACCCCT 39517  
384 erAlaProProArgLeuAlaLeuSerSerAlaLeuProGlyProProLeu 400  
|||||  
39518 CAGCCCTCTCTGGCTGGCCCTCAGCTCTGCCCTCCCTGGGCCCTCTG 39567  
401 ProAlaArgGlyHisAlaLeuLeuArgTrrPrrAlaLeuLeuCysLeuMe 417  
|||||  
39568 CCGCTCGGGGGATGCACCTGCTGGCTGGAGCGGACGCTGTGTGTGTAT 39617  
417 tValPheValSerAlaPheSerPheGlyPheGly..... 428  
|||||  
39618 GGCTTTGTGAGTGCCTTCTCTTTGGTTGGCCAGGTAAGTGGAGTT 39667  
428 ..... 428  
39668 TTCTTCAGGTGACTCTGGAGACTTCTACCCCTCTAGGGGGAAGTTGG 39717  
428 ..... 428

39718 GGACTTCCCACTGATGACCTGGCAGGGTGTCTAGTGGGCGACTAGAGGT 39767  
428 ..... 428  
39768 GGGTTGACCATGAAGCCCTCAGGGCCCTCCTACTTACCTGCGAGTTGCTCCTC 39817  
428 ..... 428  
39818 CCAGGGCCAGTGCCTAATTTTATGTTGTAATTTGTAATTTCTTCTTAAAGCC 39867  
428 ..... 428  
39868 TCCCAAGTTCATTTGAGCTTCAGCTCCCTTAAAGCCCGGATAGCTCAAA 39917  
428 ..... 428  
39918 CTGGAAGGAATTTAATGAATAGCTGAAACCAATCATCTCCAGTTCGTGCC 39967  
428 ..... 428  
39968 AAGACTGTTCCTCATCTCGGCATTGAAAGTCTCATGTCTCAGAAAACCTCT 40017  
428 ..... 428  
40018 CAGTCTTTCAGAAACTGAGACTCTACTGTAATTTCTTATAGCAGAAATGGA 40067  
428 ..... 428  
40068 GTGGGAGTTTAGTATCAGGAGCCTGACACTAGTAAGAAATCCCATTTGTTAGA 40117  
428 ..... 428  
40118 GGCTGCATGTTTGACCTGATGGTCCAGGCCCGCCAGTCCACACAGCCC 40167  
429 ..... Pro..ValThrTrrPrrLeuValLeuSerGlu 437  
40168 AGGAGCCCTCCTGCTCTCTCCACCCCTAGTACCTGGCTTGTCTCAGCGAG 40217  
438 lIeTrrProValGluIleArgGlyArgAlaPheAlaPheCysAsnSerPh 454  
|||||  
40218 ATCTACCTCTGGAGATACGAGGAGAGGCTTCCGCTTCTGCAACAGCTT 40267  
454 eAnTrrPrrAlaAlaAsnLeuPheIleSerLeuSerPheLeuAspLeuIleG 471  
|||||  
40268 CAACCTGGGCGGCCAACCTCTTCATCAGCCTCTCCTCTCCTCATCTCATTTG 40317  
471 ly 471  
40318 GT 40319  
seq\_name: gb\_htg:AL591064  
seq\_documentation\_block:  
LOCUS AL591064 200125 bp DNA HTG 03-MAY-2001  
DEFINITION Mus musculus chromosome 2 clone RP23-395E18, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in unordered pieces.  
ACCESSION AL591064  
VERSION AL591064.3 GI:13990651  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 200125)  
Sims,S.  
REFERENCE Direct Submission  
AUTHORS Submitted (02-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
TITLE CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk  
JOURNAL requests: clonerequest@sanger.ac.uk  
COMMENT On May 7, 2001 this sequence version replaced gi:13990284.  
----- Genome Center  
Center: Sanger Centre



```

/note="assembly_fragment:01520
fragment_chain:1"
190618..195019
/note="assembly_fragment:00987
fragment_chain:2"
196120..200125
/note="assembly_fragment:00065
fragment_chain:2
clone_end:sp6
vector_alde:right"
BASE COUNT 50402 a 48944 c 49953 g 49224 t 1502 others
ORIGIN

alignment_scores:
    Quality: 1678.00      Length: 549
    Ratio: 4.014          Gaps: 4
    Percent Similarity: 64.407    Percent Identity: 56.394

alignment_block:
US-09-652-292-2 x AL591064 ..
Alignment seq 1/1 to: AL591064 from: 1 to: 200125

```

US-09-032-292-2 X AL391004

[illegible]

164543 CAGGGAAGGGAGACAAGACAAACCGGGCCCTGGTGAAGCCACAGTACAGCTT 164592  
164592 eLeuAspLeuPheArgAlaArgAspAsnMetArgGlyArgThrThrValG 235  
164593 TCTGGACCTTTCAGGGCCAGGACGGATGTGGAGCCGAGCTAGTAGG 164642  
235 lYLeuGlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnValLeu 251  
164643 GGCTGGGCGCTGTGCTGTTTCAGCAGCTCAGAGCAGCCCAATGTGCTG 164692  
252 CysTyrAlaSerThrIlePheSerSerValGlyPheHisGlyGlySerSe 268  
164693 TATTACGGCTCTACCATCTTTCGCTCCCGTGGCTTCCACGGGGGCTCCTC 164742  
268 rAlaValLeuAlaSerValGlyLeuGlyAlaValLysValAlaIaThrL 285  
164743 AGCTGTGTGCTCTCTGTGGGCTTGGCACGCTGAAGCTGGCTGCCACCT 164792  
285 euThrAlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeuLeu 301  
164793 TGGTGGCCACGGGGCTGGTGACCGCTGCAGCCGCCAGAGTCTCTCTGCTC 164842  
302 AlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuValSe 318  
164843 TTTGGATGGCTCTCATGGCCCTTATCGGTCAAGCGCATAGGCGCTGGTCAG 164892  
318 rPheAlaValProMetAspSerGlyProSerCysLeuAlaValProAsnA 335  
164893 CTTTGTCTGTGCTCTGGAGCTCTGGACCCAGCTGCTGSCCACCATCCAATG 164942  
335 lathrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAspSer 351  
164943 CCAGCCAGGAGGTGGATCTGCTCGTGAAGCTCGGGTTGCTTGTGCGTTCA 164992  
352 SerLeuProIleProArgThrAsnGluAspGlnArgGluProIleLe 368  
164993 TCTCTGGCGCCAGTGTCTACACACCAANTGGGACCAAGNACAGCTGGTCT 165042  
368 uSerThrAlaLysLysThrLysProHisProArgSerGlyAspProSerA 385  
165043 GTCACTGACTGAGAGA...CCCATCCATCCAGTCATCACAGCATCCCTGG 165089  
385 laProProArgLeuAlaLeuSerSerAlaLeuProGlyProProLeuPro 401  
165090 GACCA.....GTTCTGAACACTGCCCTCCCGAGTCCACCCAGTCCC 165130  
402 AlaArgGlyHisAlaLeuLeuArgTrpThrAlaLeuLeuCysLeuMetVa 418  
165131 ATCCTTGAGCATACCCCTGCTGTGCTGCTCGGCTGGTTTCCATGATGGT 165180  
418 lPheValSerAlaPheSerPheGlyPheGlyPro.Val..... 430  
165181 CTACGTGAGCGCCTCTCTCGTTGGCTTTGGACAGGTAGGAGGGCTTATC 165230  
430 ..... 430  
165231 GGTGAGGGAATCTAGAGCCTTCTGTCCCTTTTCAGTGGGCTCTTATAGTC 165280  
430 ..... 430  
165281 TAGGGACACTCAGAGCTTTAGTGATGGCAGCGGAGGCTCCAGGGTTCCTT 165330  
430 ..... 430  
165331 GCTTATATCAAGGGGCTCCCGCCATAAGTTTAAATCCAACACCATCTGCAA 165380  
430 ..... 430  
165381 GTTTTAAAGCAACCTCCCTGCTCCCGTGTGGGACCAACATTAAA 165430  
430 ..... 430

165431	CTCCAGGTCCTCTCAAAACCTTAGAACCTATCCCTGAGTGGTTCCATTAGAA	165480
430	.....	430
165481	GGCTGGCTTGAAGAGAAATGTAACAGTCAGATAAAACCAGTCATACCA	165530
430	.....	430
165531	GTTTGCCCTGGGGGGGGGGGCCCTTCTGCTTGGCATGCAATGTTCTGTG	165580
430	.....	430
165581	TCCAGGAACCTGAGACTGCCCTATTTCATGGAGTTCCTCCAGGGGCACA	165630
430	.....	430
165631	GAAGTTTGGCATCTGAGAGAAATCTGGGGCCAAAGAGCATGACACAGGTCTC	165680
430	.....	430
165681	AGTTAAACGAGGGTGGGAGGTGGCCCTGTCCAGTTTCAGTGGGTGCCCAAGATC	165730
431	.....ThrTrrpLeuValLeuSerGluIleTyrp	440
165731	TCTCTTGTCTTCGGCTCCAGTAACCTGGTTGGTCTCTCAGTCAGATCTACC	165780
440	rovalGluIleAArgGlyArgAlaPheAlaPheCysAsnSerPheAsnTrrp	456
165781	CAGCGGAGATCCGAGGGAGAGCCCTTCGGCTTCTGCAGCAGCTTCAACATGG	165830
457	AlaAlaAsnLeuPheIleSerLeuSerPheLeuAspLeuIleGly	471
165831	GCAGCTAACCTCTTTATCAGGCTCTCCTTCTCGACCTGATCGGGT	165875

seq\_name: gb\_htq:AL591805

seq\_documentation\_block:

LOCUS	AL591805	243075 bp	DNA	HTG	22-JUL-2001
DEFINITION	Mus musculus chromosome 2 clone RP23-90N15, *** SEQUENCING IN PROGRESS ***, in unordered pieces.				

ACCESSION AL591805

RECESSION  
VERSION  
AL591805.9  
AL591803  
GI:14626203

**KEYWORDS**

**SOURCE**  
house mouse.

ORGANISM    Mus musculus

Eukaryota; Metazoa; Cho

Mammalia; Eutheria; Rodentia

REFERENCE 1 (bases 1 to 243075)

**AUTHORS** Sims, S.

TITLE	Direct Submission
TOURNAMENT	Category: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, AA, AB, AC, AD, AE, AF, AG, AH, AI, AJ, AK, AL, AM, AN, AO, AP, AQ, AR, AS, AT, AU, AV, AW, AX, AY, AZ, BA, BB, BC, BD, BE, BF, BG, BH, BI, BJ, BK, BL, BM, BN, BO, BP, BQ, BR, BS, BT, BU, BV, BW, BX, BY, BZ, CA, CB, CC, CD, CE, CF, CG, CH, CI, CJ, CK, CL, CM, CN, CO, CP, CQ, CR, CS, CT, CU, CV, CW, CX, CY, CZ, DA, DB, DC, DD, DE, DF, DG, DH, DI, DJ, DK, DL, DM, DN, DO, DP, DQ, DR, DS, DT, DU, DV, DW, DX, DY, DZ, EA, EB, EC, ED, EE, EF, EG, EH, EI, EJ, EK, EL, EM, EN, EO, EP, EQ, ER, ES, ET, EU, EV, EW, EX, EY, EZ, FA, FB, FC, FD, FE, FF, FG, FH, FI, FJ, FK, FL, FM, FN, FO, FP, FQ, FR, FS, FT, FU, FV, FW, FX, FY, FZ, GA, GB, GC, GD, GE, GF, GG, GH, GI, GJ, GK, GL, GM, GN, GO, GP, GQ, GR, GS, GT, GU, GV, GW, GX, GY, GZ, HA, HB, HC, HD, HE, HF, HG, HH, HI, HJ, HK, HL, HM, HN, HO, HP, HQ, HR, HS, HT, HU, HV, HW, HX, HY, HZ, IA, IB, IC, ID, IE, IF, IG, IH, II, IJ, IK, IL, IM, IN, IO, IP, IQ, IR, IS, IT, IU, IV, IW, IX, IY, IZ, JA, JB, JC, JD, JE, JF, JG, JH, JI, JJ, JK, JL, JM, JN, JO, JP, JQ, JR, JS, JT, JU, JV, JW, JX, JY, JZ, KA, KB, KC, KD, KE, KF, KG, KH, KI, KJ, KK, KL, KM, KN, KO, KP, KQ, KR, KS, KT, KU, KV, KW, KX, KY, KZ, LA, LB, LC, LD, LE, LF, LG, LH, LI, LJ, LK, LL, LM, LN, LO, LP, LQ, LR, LS, LT, LU, LV, LW, LX, LY, LZ, MA, MB, MC, MD, ME, MF, MG, MH, MI, MJ, MK, ML, MM, MN, MO, MP, MQ, MR, MS, MT, MU, MV, MW, MX, MY, MZ, NA, NB, NC, ND, NE, NF, NG, NH, NI, NJ, NK, NL, NM, NN, NO, NP, NQ, NR, NS, NT, NU, NV, NW, NX, NY, NZ, OA, OB, OC, OD, OE, OF, OG, OH, OI, OJ, OK, OL, OM, ON, OO, OP, OQ, OR, OS, OT, OU, OV, OW, OX, OY, OZ, PA, PB, PC, PD, PE, PF, PG, PH, PI, PJ, PK, PL, PM, PN, PO, PP, PQ, PR, PS, PT, PU, PV, PW, PX, PY, PZ, QA, QB, QC, QD, QE, QF, QG, QH, QI, QJ, QK, QL, QM, QN, QO, QP, QQ, QR, QS, QT, QU, QV, QW, QX, QY, QZ, RA, RB, RC, RD, RE, RF, RG, RH, RI, RJ, RK, RL, RM, RN, RO, RP, RQ, RR, RS, RT, RU, RV, RW, RX, RY, RZ, SA, SB, SC, SD, SE, SF, SG, SH, SI, SJ, SK, SL, SM, SN, SO, SP, SQ, SR, SS, ST, SU, SV, SW, SX, SY, SZ, TA, TB, TC, TD, TE, TF, TG, TH, TI, TJ, TK, TL, TM, TN, TO, TP, TQ, TR, TS, TT, TU, TV, TW, TX, TY, TZ, UA, UB, UC, UD, UE, UF, UG, UH, UI, UJ, UK, UL, UM, UN, UO, UP, UQ, UR, US, UT, UU, UV, UW, UX, UY, UZ, VA, VB, VC, VD, VE, VF, VG, VH, VI, VJ, VK, VL, VM, VN, VO, VP, VQ, VR, VS, VT, VU, VV, VW, VX, VY, VZ, WA, WB, WC, WD, WE, WF, WG, WH, WI, WJ, WK, WL, WM, WN, WO, WP, WQ, WR, WS, WT, WU, WV, WW, WX, WY, WZ, XA, XB, XC, XD, XE, XF, XG, XH, XI, XJ, XK, XL, XM, XN, XO, XP, XQ, XR, XS, XT, XU, XV, XW, XX, XY, XZ, YA, YB, YC, YD, YE, YF, YG, YH, YI, YJ, YK, YL, YM, YN, YO, YP, YQ, YR, YS, YT, YU, YV, YW, YX, YY, YZ, ZA, ZB, ZC, ZD, ZE, ZF, ZG, ZH, ZI, ZJ, ZK, ZL, ZM, ZN, ZO, ZP, ZQ, ZR, ZS, ZT, ZU, ZV, ZW, ZX, ZY, ZZ

JOURNAL Submitted (21-JUL-2001)

CB10 1SA, UK. E-mail en

COMMENTS: On Jul 7 2001 4:45 PM requests: clonerequest@

COMMENT On Jul 7, 2001 this seq

Genome C

Center: Sanger Centre  
Center code: SC

Center code: SC  
Web site: <http://www.scc.edu>

Web site: <http://www.sai>  
Contact: [humquerv@sai](mailto:humquerv@sai)

Contact: humquerry@esange.com  
----- project -----

Center project name: bmc

center project name: BM  
----- Summary S

Assembly program: XGAP4

Sequencing vector: pBR322

**Chemistry:** Dye-terminator

Consensus quality: 23854

Consensus quality: 23973

Consensus quality: 24056

Insert size: 241575; sum

Insert size: 239393; 4.9

Quality coverage: 8.35x

-----  
coverage: 8.51x in Q20 R

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84



302 AlaGlyCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuValSer 318  
211086 TTTGATGCGCTCTCATGCGCTTATCGGTACGGCATAGGCTGGTTCAG 211037  
318 rPheAlaValProMetAspSerGlyProSerCysLeuAlaValProAsnA 335  
211036 CTTTGTCTGTCCCTGGGACTCTGGACCCAGCTGCGCTGGCCACATCCAAATG 210987  
335 laThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnAspSer 351  
210986 CCAGCCAGCAGGTGGATCTGCTGCTGAAGCTCGGGTTTGTCTGGCTTCA 210937  
352 SerLeuProProlIleProArgThrAsnGluAspGlnArgGluProlIle 368  
210936 TCTCTGGCGCCAGTCTACACACCAATGGGACCAAGGACAGCTGCTCT 210887  
368 userThrAlaLysIleThrLysProHisProArgSerGlyAspProSerA 385  
210886 GTCACTCACTGAGAGA...CCATCCATCCAGTCAATCAGCATCCCTGG 210840  
385 laProProArgLeuAlaLeuSerSerAlaLeuProGlyProLeuPro 401  
210839 GACCA...GTTCTGAACACTGCTCCCAAGTTCACCACAGTCCC 210799  
402 AlaArgGlyHisAlaLeuLeuArgTrpThrAlaLeuLeuCysLeuMetVa 418  
210798 ATCCTTCAGCATACCCCTGCTGTGCTGCTGCGCTGGTTTGCATGATG 210749  
418 lPheValSerAlaPheSerPheGlyPheGlyPro.Val 430  
210748 CTAGTGAGCGCCCTCTCGTGGCTTTGGACCAAGTACGGAGGCTTATC 210699  
430 ..... 430  
210698 GGGTAGGGGAATCTAGAGCCTTCTGCTCCCTTTCAGTGGGCCTTCTATGTC 210649  
430 ..... 430  
210648 TAGGGACACTCGAGCTTTAGTGATGGCAGCGAGGCTCCAGGGTTCTCTT 210599  
430 ..... 430  
210598 GCTTATATCAAGGGGCTCCCCCATAGTTTAAATCCAAACCATCTGCAA 210549  
430 ..... 430  
210548 GTTTTAAAGCCAACCCCTCCCTCTGCCCGTGTGGGCACCAACATTTAA 210499  
430 ..... 430  
210498 CTCACGCTCTTTCAACCTTAGACCTATCCCTGAGTGGTTCCCATAGAA 210449  
430 ..... 430  
210448 GGGCTGGCTTGAAGAGAACTGTAACGAGTCAAGATAAAACCATCATACCA 210399  
430 ..... 430  
210398 GTTTGCTGGGGGGGGGGCCCTTCTGCTTGCATCGAATGTCTTGTG 210349  
430 ..... 430  
210348 TCCAGGAACCTGAGACTCGCGCTATTCTATGAGTTCTCCAGGGGGCACA 210299  
430 ..... 430  
210298 GAAGTTGTGATCTGAGAGAACTGGGGCCAAAGCATGACACAGGCTC 210249  
430 ..... 430  
210248 AGTTAACGAGGTGGGAGGTGGCTGTCCAGTTCTATGGGTGGCCAGAAATC 210199  
431 .....ThrTrpLeuValLeuSerGluIleTrp 440

210198	TCCTCTGCTCGCGCTCCAGTAACCTGGTGGTCTCAGTGAGATCTACC	210149
440	roValGluileAtGlyArGAlapheAlapheCysAsnSerPheAsnTrp	456
210148	CAGCGGACATCGAGGAGAGCCTTCCGCTTCTGCAGCAGCTTCAACTGG	210099
457	AlaAlaAsnLeuPheLeSerLeuSerPheLeuAspLeuileGly	471
210098	GCAGCTAAACCTCTTTATCAGCCCTCTCTCTCTCGACCTGATCGGT	210054
seq_name: gb_pr:AB056798		
seq_documentation_block:		
LOCUS	AB056798	4521 bp mRNA PRI 14-MAR-2001
DEFINITION	Macaca fascicularis brain cDNA clone:Of1A-11110, full insert sequence.	
ACCESSION	AB056798	
VERSION	AB056798.1	GI:13365896
KEYWORDS	fis (full insert sequence); oligo capping.	
SOURCE	Macaca fascicularis adult male frontal lobe left cDNA to mRNA, clone_lib:macaque brain cDNA library Of1A clone:Of1A-11110.	
ORGANISM	Macaca fascicularis	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae; Cercopitheciinae; Macaca.	
AUTHORS	1 (sites)	
TITLE	Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.	
JOURNAL	Isolation of full-length cDNA clones from macaque brain cDNA libraries	
REFERENCE	2 (bases 1 to 4521)	
AUTHORS	Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.	
TITLE	Direct Submission	
JOURNAL	Submitted (09-MAR-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan	
COMMENT	(E-mail: khashienh.go.jp, URL: http://www.nih.go.jp/yoken/genebank/, Tel: 81-3-5285-1111(ex.2120), Fax: 81-3-5285-1181)	
Vector:	pME18S-FL3 (Acc.No. AB009864)	
R. Site1:	DrIII (CACTGTGTG)	
R. Site2:	DrIII (CACCATGTG)	
Description:	1st strand cDNA was primed with an oligo(dt) primer [ATGTGGCTTTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DrIII sites of pME18S-FL3. XhoI sites just outside the DrIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing { 5' end primer [CTTCTGCTCTTAAAGTCGG] ; 3' end primer [CGACCTGCAGCAGACA] }.	

FEATURES  
SOURCE

```

L: 45321
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="Of1A-11110"
/sex="male"
/tissue_type="frontal lobe left"
/clone_lib="macaque brain cDNA library Of1A"
/dev_stage="adult"
/node_host="TOP10; Vector: pME18S-FL3 (Acc.No. AB009864);
R. Site1: DraIII (CACTGTGTG); R. Site2: DraIII
(CACCATGTG)"
157..2022
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB93322.1"
/db_xref="gi:13365897"
/translation="MVPEVTEGPNLLNOKGTAVETEGSVRAGSRHPPDWARCGCMGPT

```

FLSSVTAANVSGLLVGYELCIIIGALLQIKTLTLTSCHEQEMVYSSLLIGALLASLTG6G  
VLIDTGRRTATILSSCLLIGLGSVLISUSYTVLIVRGTAIGVSLVSLGICLACTVIA  
ELDQHRRLGLLVSLLGKLVIGLISVLSNAYEAFNPHGWKMGFLVPIGLIOALAM  
VYLPSPRFLVMKQGEAKSLGRLASDATEELTVKSSLKDEYQVSGVDFLTSK  
DNMRITMGLTLVFFVQIQGNRLIFYASTVLKSVGFQCSNEAASLQSGVGVYVVIS  
TIPATLLVDHVGSKTLFLCIGSVMAASLVTMGVLNLIHMNFNTSASLLNAVLSHTE  
SVTYGPNGLSASNTLRIDPKGTAHSRSSLMLRNDVDKRGETTSSASLLNAVLSHTE  
YQIYTDGDPVAPFLKSLASLLVYVYANFSGIGCPMPWLVLSEIFPGGIGRAMALITS  
SMWGNLILSLTFLTVDIGUPVCFYITIMSLASLFLVVMFIPQTRGCSLEQISM  
ELAKVNYVNNKTCFMSHQEELPVQPKRKPQEQLELCKNLKGRGQSGLSPEN"  
1340 a 943 c 920 g 1318 t  
BASE COUNT  
ORIGIN

```

alignment_scores:
  Quality: 932.00      Length: 551
  Ratio: 2.596        Gaps: 7
  Percent Similarity: 65.154  Percent Identity: 39.201

alignment_block:
  US-09-652-292-2 x AB056798  ..

  Align seg 1/1 to: AB056798 from: 1 to: 4521

10  LeuCysA5erValSerLeuLeuGlyGlyLeuThrPheGlyTyrGluLe 26
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
292  CTGTCATCTGTCATCCGCCGTGACAGTGGCTCTGGTGGTTATGAAC 341
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
26  uAlaValIleSerGlyAlaLeuLeuProLeuGlnLeuAspPheGlyLeu 43
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
342  TGGGATCATCTCTGGGCTCTCTTCAGATCAAAACCTTATTAAACCTGA 391
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
43  erCysLeuGlnGluPheLeuValGlySerLeuLeuGlyAlaLeu 59
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
392  GCTGCCATGACGAGGAATGGTTGTGAGTCCCTCTCATGGAGCCCTC 441
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
60  LeuAlaSerLeuValGlyGlyPheLeuIleAspCysTyrGlyArgLys 76
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
442  CTCGCCCTCACTCACCGGAGGGTCTCTATAGCAGGTATGGAAGAGGAC 491
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
76  nAlaIleLeuGlySerAsnLeuValLeuLeuAlaGlySerLeuThrLeu 93
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
492  CGCAATCATCTTGTCTATCCTTCCTGCTGGACTCGGAAGCTTAGTCTTGA 541
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
93  lyLeuAlaGlySerLeuAlaTrpLeuValLeuGlyArgAlaValGly 109
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
542  TCCTCAGTTTATCCTACACGGTCTTATAGTGGAGCGCATTGCCATAGG 591
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
110  PheAlaIleSerLeuSerSerMetalCysCysIleTyrValSerGluLe 126
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
592  GTCTCCATCTCCCTCTCTCCATTTGCCACTTGTGTTTACATCGCAGAGAT 641
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
126  uValGlyProArgGlnArgGlyValLeuValSerLeuTyrGluAlaGly 143
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
642  TGTCTCTCAACACAGAAGCGCTTCTTGTGCTACTGAATGAGCTGATGA 691
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
143  leThrValGlyIleLeuLeuSerTyrAlaLeuAsnTyrAlaLeuGly 159
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
692  TTGTCTATCGGCATCTTCTGCTTATATTTCAAATTACGCATTTGCCAAT 741
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
160  ThrProTrpClyTrpArgHisMetPheGlyTrpAlaThrAlaProAla 176
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
742  GTTTTCCATGGCTGGAAGTACATGTTTTGGTCTTGATTCCTTGGGAAT 791
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
176  lleuGlnSerLeuSerLeuPheLeuProAlaGlyThrAspGluThrA 193
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
792  TTTGCAAGCAATTCGAATGATTATTCTTCTCCCAAGC..... 828
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
193  laThrHisLysAspLeuIleProLeuGlnGly.....GlyGluAlaPro 207
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
829  .....CCTCGGTTTCTGTGATGAAGAGACAAAGAGGAGGCTCTGCTAGC 870
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:
208  LysLeuGlyProGlyArgProArg..... 215
   ||| :||: :||: :||: :||: :||: :||: :||: :||: :||:

```

1714 CTGACATTTTTCAGCTGAACGATCTTATGGCCCTGCCATGGGTGTGCTT 1763

480 uLeuTyrGlyLeuThrAlaValLeuGlyLeuGlyPheIleTyrLeuPheV 497

1764 TATATATACATCATGAGCTAGACCTCTGCTTTTGTGTATGTTTA 1813

497 alProGluThrLysGlyGlnSerLeuAlaGluIleAspGlnGlnPheGln 513

1814 TACCTGACACAAGGGTCTCTTTGGACAACAATATCAATGAGCTAGCA 1863

514 Lys 514

1864 AAA 1866

seq\_name: gb\_pr:AL449363

seq\_documentation\_block:

LOCUS AL449363 141289 bp DNA PRI 19-APR-2001  
DEFINITION Human DNA sequence from clone CTA-444M12 on chromosome 6, complete sequence.

ACCESSION AL449363

VERSION 1 (bases 1 to 141289)

KEYWORDS HTG.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 141289)

Direct Submission

Submitted (19-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

Requests: clonerequest@sanger.ac.uk

On Apr 21, 2001 this sequence version replaced gi:13446493.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em: EMBL; Sw:

SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6

CTA-444M12 is from the human BAC library described in U-J. Kim et

al (1996) Genomics 34, 213-218.

VECTOR: pBelOBAC11

IMPORTANT: This sequence is not the entire insert of clone

CTA-444M12 it may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true left end of clone RPI-188K17 is at 141190 in this

sequence. The true right end of clone RPI-73H22 is at 100 in this

# FEATURES

## source

Location/Qualifiers

1. 141289

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="6"

/clone\_lib="CIT978SK-A2"

69075. 69084

## misc\_feature

BASE COUNT 41833 a 29135 c 29528 g 40793 t  
ORIGIN  
/note="1329bp of tn10 transposon (J01829) removed here."

alignment\_scores:  
Quality: 727.00 Length: 472  
Ratio: 2.440 Gaps: 7  
Percent Similarity: 63.136 Percent Identity: 37.500

alignment\_block:

US-09-652-292-2 x AL449363/rev

Align seg 1/1 to reverse of: AL449363 from: 1 to: 141289

10 LeuCysAlaSerValSerLeuLeuGlyLeuThrPheGlyTyrGluLeu 26

23675 CTGTCACTGTCACCTGCTGTCAGTGCCTCTCTGGTGGGTATGAAC 23626

26 uAlaValIleSerGlyAlaLeuProLeuGlnLeuAspPheGlyLeu 43

23625 TGGGATCATCTCTGGGGCTTCTTCAGATCAAAACCTTATTAGCCCTGA 23576

43 erCysLeuGluGlnGluPheLeuValGlySerLeuLeuGlyAlaLeu 59

23575 GCTCCCATGACGAGAAATGGTGTGAGTCCCTCGTCATTTGGAGCCCTC 23526

60 LeuAlaSerLeuValGlyGlyPheLeuIleAspCysTyrGlyArgLysG 76

23525 CTTGCCCTCACTACCGGGGCTCTGTATAGACAGATATGAAGAGGAC 23476

76 nAlaIleLeuGlySerAsnLeuValLeuAlaGlySerLeuThrLeu 93

23475 AGCAATCATCTGTTCCTGCTGCTGGACTCGGAAGCTTAGCTTGA 23426

93 lyLeuAlaGlySerLeuAlaTriLeuValLeuGlyArgAlaValValG 109

23425 TCCTCAGTTTATCTACACGGTCTTATAGTGGACGCGATGCCATAGG 23376

110 PheAlaIleSerLeuSerSerMetAlaCysCysIleTyrValSerGlu 126

23375 GTCTCCATCTCCCTCTCTTCATTTGCCACTTGTGTTCATCGCAGAG 23326

126 uValGlyProArgGlnArgGlyValLeuValSerLeuTyrGluAlaG 143

23325 TGCTCCTCAACAGAGAGCGCTTCTGTGCTACTGAAGAGCTGATGA 23276

143 leThrValGlyIleLeuSerTyrAlaLeuAsnTyrAlaLeuAlaGly 159

23275 TTGTATCGGCATCTCTTCTGCTCATATATTTCAAAATACGCAATTGCCA 23226

160 ThrProTyrGlyTyrArgHisMetPheGlyTyrAlaThrAlaProAla 176

23225 GTTTTCATGGCTGGAGATGACATGTTTGGTCTGTGATTCCTCTGGGAG 23176

176 lLeuGlnSerLeuSerLeuPheLeuProAlaGlyThrAspGluThr 193

23175 TTTTGAAGCAATGCAATGATTTTCTCTCTCCCAAGC. 23139

193 laThrHisIleAspLeuIleProLeuGlnGly 207

23138 .....CCTCGGTTTCTGGTGTATGAAGGACAAGAGGAGCTGCTAGC 23097

208 LysLeuGlyProGlyArgProArg 215

23096 AAGTTT...CTTGAAGGTTTAAAGGCACTCTCAGATACAACTGAGGA 23050

216 .....TyrSerPheLeu 220

23049 CACTGTGATCAAAATCTCCCTGAAAGATGAATATCAGTACAGTTTGGG 23000

220 spleuPheArgAlaArgAspAsnMetArgGlyArgThrThrValGlyLeu 236

```

22999 ATCTGTTTCCTTCAAAAGACAACATCGGACCGCAATAATGATAGGACTA 22950
237 GlyLeuValLeuPheGlnGlnLeuThrGlyGlnProAsnValLeuCys 253
22949 ACATAGTATTTTGTACAAATCACTGGCGCAACCAACATATTTGTTCTA 22900
253 rAlaSerThrIlePheSerSerValGlyPheHisGlyGlySerSerAlaVal 270
22899 TGCATCAACTGTTTGAAGTCACTGGATTTCAAAGCAATGAGGCGAGCTA 22850
270 aLeuAlaSerValGlyLeuGlyAlaValValValValAlaAlaThrLeuThr 286
22849 GCCTCGCCCTCACTGGGTTGGAGTCTCAAGTCTATTAGCACCATCCCT 22800
287 AlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeuLeuAlaGln 303
22799 GCCACTCTTCTTGTAGACCATGTCGCGAGCAACCAACATTCCTTCGATTGG 22750
303 yCysAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuValSerPheA 320
22749 CTCTCTGTGATGGCAGCTTCTGTTGGTGCACCATGGGCATCGTAATCTCA 22700
320 laValProMetAsp..... 324
22699 ACATCCACATGAATCTCACCCATATCTGCAGAGCCACAAATTCATCAAC 22650
325 .....SerGlyProSerCysLeuAlaValPr 333
22649 CAGTCTCTGGATGAGTCTGTGATTTATGATGACCAACCTGTCAACCAA 22600
333 oAsnAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnA 350
22599 CAACAATACTCTCAGAGACCACTTCAAGGGGATTTCTTCCCATAGCAGAA 22550
350 spSerSerLeuPro.....ProIleProArgThrAsnGlyAspGln 363
22549 GCTCACTATGCCCTCGAGAAATGATGTGATAGAGAGGGAGACCAACC 22500
364 ArgGluProIleLeuSerThrAlaLysLysThrLysProHisProArgse 380
22499 TCAGCATCTCTGCTTAATGCTGATTAAGCCACACTGAATACCATAGT 22450
380 rGlyAspProSerAlaProProArgLeuAlaLeuSerSerAlaLeuProG 397
22449 CACAGACCTGGGAGCGTCCCA..... 22428
397 lyProProLeuProAlaArgGlyHisAlaLeuLeuArgTTPThrAlaLeu 413
22427 .....GCTTTTGTGAATGGCTGCTCTTA 22404
414 LeuCysLeuMetValPheValSerAlaPheSerPheGlyPheGlyProva 430
22403 GCCAGCTTGTCTTTATGTTGCTGCTTTTCAATTGCTAGGACCAAG 22354
430 lThrTpLeuValLeu 435
22353 TAAGTACTTATTCTT 22338
seq_name: gb_hhg:AC023293
seq_documentation_block:
LOCUS AC023293 163569 bp DNA HTG 01-MAR-2000
DEFINITION Homo sapiens clone RP11-17F18, WORKING DRAFT SEQUENCE, 17 unordered
pieces.
ACCESSION AC023293
VERSION AC023293.2 GI:7138735
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 163569)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Homo sapiens, clone RP11-17F18  
Unpublished  
2 (bases 1 to 163569)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavsky,L.,  
Boukhgalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collino,M., Cooke,P.,  
DeArrellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,  
Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Illey,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,  
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,  
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,  
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,  
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,  
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,  
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Willson,B.,  
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
Zody,M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 1, 2000 this sequence version replaced gi:6960435.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3635

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 153600 bases at least Q40

Consensus quality: 158799 bases at least Q30

Consensus quality: 160722 bases at least Q20

Insert size: 176000; agarose-fp

Insert size: 161969; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; agarose-fp

Quality coverage: 4.3 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1467: contig of 1467 bp in length  
\* 1468 1567: gap of 100 bp  
\* 1568 3349: contig of 1782 bp in length  
\* 3350 3449: gap of 100 bp  
\* 3450 6675: contig of 3226 bp in length  
\* 6676 6775: gap of 100 bp  
\* 6776 9516: contig of 2741 bp in length  
\* 9517 9616: gap of 100 bp  
\* 9617 13613: contig of 3997 bp in length  
\* 13614 13713: gap of 100 bp  
\* 13714 17758: contig of 4045 bp in length  
\* 17759 17858: gap of 100 bp in length  
\* 17859 22031: contig of 4173 bp in length  
\* 22032 22131: gap of 100 bp



```

US-09-652-292-2 x AC023293      .
Align seg 1/1 to: AC023293 from: 1 to: 163569

10 LeuCysAlaSerValSerLeuLeuGlyGlyLeuThrPheGlyTyrGluLeu 26
||||| ::: :::: ||||| :::: ||||| :||| ||||| |||||
54849 CTTGCATCTGTCACTGCTGCTGTCAAGTGCGCTCCTGGTATTGAAC 54898
26 uAlaValIleSerGlyAlaLeuLeuProLeuInLeuAspPheGlyLeu 43
||||| :||| ||||| :||| :||| :||| :||| :||| :|||
54899 TGGGATCATCTCTGGGGCTCTTCTTCAGATCAAAACCCTTAATGCCCTGA 54948
43 eCysLeuGluGlnGluPheLeuValGlySerLeuLeuLeuGlyAlaLeu 59
||||| :||| :||| :||| :||| :||| :||| :||| :|||
54949 GCTGCCATGAGCAGGAAATGGTTGTGAGCTCCCTCGTCATTGGAGCCCTC 54998
60 LeuAlaSerLeuValGlyGlyPheLeuIleAspCysTyrGlyArgLysGI 76
||||| :||| :||| :||| :||| :||| :||| :||| :|||
54999 CTTCCTCACTCACCGGAGGGTCCGTGATAGACAGATATGGAAGAGGAC 55048
76 nAlaIleLeuGlySerAsnLeuValLeuAlaGlySerLeuThrLeuG 93
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55049 AGCAATCATCTGTCACTCGCTGCTGTGGAGTCGGAAGCTTAGTCTTGA 55098
93 LysLeuAlaGlySerLeuAlaTrpLeuValLeuGlyArgAlaValGly 109
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55099 TCCTCAGTTTATCTACACGGTTCCTTATAGTGGAGCGCATTGCATAGG 55148
110 PheAlaIleSerLeuSerMetAlaCysCysIleTyrValSerGluLeu 126
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55149 GTCTCCATCTCCTCTCCATTGGCCTTGTGTTACATCGCAGAGAT 55198
126 uValGlyProArgGlnArgGlyValLeuValSerLeuTyrGluAlaGlyI 143
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55199 TGCTCCTCAACAACAAAAGGCTCTCTGTGCACATGAATGAGCTGATGA 55248
143 IeThrValGlyIleLeuLeuSerTyrAlaLeuAsnTyrAlaLeuAlaGly 159
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55249 TTGTATCGGCATCTTCTTCGCTATATTTCAAATACGCATTTGGCAAT 55298
160 ThrProTrpGlyTrpArgHisMetPheGlyTrpAlaThrAlaProAlaVa 176
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55299 GTTTTCCATGGCTGGAAGTACATGTTTGGTCTGTGATCCCTTTGGAGT 55348
176 IleuGlnSerLeuSerLeuPheLeuProAlaGlyThrAspGluThrA 193
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55349 TTTGCAAGCAATTGCAATCTATTCTTCTCCTCCAAGC..... 55385
193 IaThrHisLysAspLeulleProLeuIngLy.....GlyGluAlaPro 207
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55386 .....CCTCGGTTCTTGTFGTATGAAGGACAGAGGAGGTGCTAGC 55427
208 LysLeuGlyProGlyArgProArg..... 215
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55428 AAGGTT...CTTGGAGGTGTAAGACACTCTCAGATACAACTGAGGA 55474
216 .....TyrSerPheLeuA 220
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55475 CACTGTGATCAAAATCCTCCCTGAAAGATGAATATCAGTACAGTTTTGG 55524
220 spLeuPheArgAlaArgAspAsnMetArgGlyArgThrThrValGlyLeu 236
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55525 ATCTGTTTCGTTCAAAAGAACATCGCGACCCGGAATAATGATAGGACTA 55574
237 GlyLeuValLeuPheGlnInLeuThrGlyGlnProAsnValLeuCystY 253
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55575 ACACGTAGTATTTTGTACAAATCACTCGCCCAACCAACATATTCCTTA 55624
253 rAlaSerThrIlePheSerSerValGlyPheHisGlyGlySerSerAlav 270
||||| :||| :||| :||| :||| :||| :||| :||| :|||
55625 TGCATCAACTGTTTGAAGTCAGTTGGATTTCAAAAGCAATGAGGACGCTA 55674
270 alLeuAlaSerValGlyLeuGlyAlaValLysValAlaAlaThrLeuThr 286

```

```

155238 183568: contig of 28332 bp in length
Location/Qualifiers
1. 163569
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="RP11-17F18"
  /clone_lib="RPC1-11 Human Male BAC"
1. 1467
  /note="assembly_fragment"
1568. 3349
  /note="assembly_fragment"
3450. 6675
  /note="assembly_fragment"
6776. 9516
  /note="assembly_fragment"
9617. 13613
  /note="assembly_fragment"
13714. 17758
  /note="assembly_fragment"
17859. 22031
  /note="assembly_fragment"
22132. 26657
  /note="assembly_fragment"
clone_end:SP6
vector_side:right
26758. 34409
  /note="assembly_fragment"
34510. 42217
  /note="assembly_fragment"
42318. 53921
  /note="assembly_fragment"
54022. 64375
  /note="assembly_fragment"
64476. 79997
  /note="assembly_fragment"
80098. 96453
  /note="assembly_fragment"
96554. 114686
  /note="assembly_fragment"
114787. 135137
  /note="assembly_fragment"
clone_end:77
vector_side:left
135238. 163569
  /note="assembly_fragment"
BASE COUNT 49345 a 32921 c 31208 g 48493 t 1602 others
ORIGIN

alignment_scores:
  Quality: 724.00      Length: 472
  Ratio: 2.430        Gaps: 7
  Percent Similarity: 63.136  Percent Identity: 37.288

alignment_block:

```

55675 GCTTCGCTCCACTGGGTGGAGTCTCAGGTCAATTAGCACCATCCCT 55724  
 287 AlaMetGlyLeuValAspArgAlaGlyArgAlaLeuLeuAlaGln 303  
 55725 GCCACTCTCTGTAGACCATCTCGCGACCAACATCTCTGCATTCG 55774  
 303 YCYaAlaLeuMetAlaLeuSerValSerGlyIleGlyLeuValSerPheA 320  
 55775 CTCCTCTGTATGGCAGCTTCGTGTGTGACCATGGCATCGTAAATCTCA 55824  
 320 laValPrometAsp..... 324  
 55825 ACATCCACATGAACCTTACCCTATCTGCAGAACCCACAACTCTATCAAC 55874  
 325 .....SerGlyProSerCysLeuAlaValPr 333  
 55875 CAGTCCTTGATGAGTCTGTGATTATGACACAGGAACCTGTCAACCAA 55924  
 333 oAsnAlaThrGlyGlnThrGlyLeuProGlyAspSerGlyLeuLeuGlnA 350  
 55925 CAACAATACTCTCAGACACCATCTCAAGGGATTCTTCCCATAGCAGAA 55974  
 350 spSerSerLeuPro.....ProIleProArgThrAsnGluaspGln 363  
 55975 GCTCAGTCATCGCCCTCAGAAATGATCTGCTAAGAGAGGGAGACGACC 56024  
 364 ArgGluProIleLeuSerThrAlaLysLysThrLysProHisProArgSe 380  
 56025 TCAGCATCTCTGTAATGCTGGATTGACGACACATGAATACAGATAGT 56074  
 380 rGlyAspProSerAlaProProArgLeuAlaLeuSerSerAlaLeuProG 397  
 56075 CACAGACCTGGGACGTCCTCA..... 56096  
 397 lyProProProAlaArgGlyHisAlaLeuLeuArgTrpThrAlaLeu 413  
 56097 .....GCTTTTGAATGGCTGCTCTTA 56120  
 414 LeuCysLeuMetValPheValSerAlaPheSerPheClyPheClyProVa 430  
 56121 GCCAGCTTGCTGTATGTGCTCTTTTCAATGGCTAGGACCAAG 56170  
 430 lThrTrpLeuValLeu 435  
 56171 TAAGTACTTTATCTT 56186  
 seq\_name: gb\_ba:BS292954  
 seq\_documentation\_block:  
 LOCUS BS292954 8189 bp DNA BCT 24-JUN-1998  
 DEFINITION B. subtilis yws[A,B,C,D,E,F,G] and gerBC genes.  
 ACCESSION Z92954  
 VERSION Z92954.1 GI:1894764  
 KEYWORDS gerBC gene; ywsc gene; ywta gene; ywtB gene; ywtC gene; ywtD gene; ywte gene; ywtf gene; ywtG gene.  
 SOURCE Bacillus subtilis.  
 ORGANISM Bacillus subtilis  
 Bacteria; Firmicutes; Bacillus/Clostridium group;  
 Bacillus/Staphylococcus group; Bacillus.  
 1 (bases 1 to 8189)  
 PRESCAN,E., Moszer,I., Boursier,L., Cruz,H., De La Fuente,V.,  
 Hullo,M.F., Lelong,C., Schleich,S., Sekowska,A., Song,B.H.,  
 Villani,G., Kunst,P., Danchin,A. and Glaser,P.  
 The Bacillus subtilis genome from gerBC (311 degrees) to l1cr (334 degrees)  
 JOURNAL Microbiology 143 (Pt 10), 3313-3328 (1997)  
 MEDLINE 98015417  
 REFERENCE 2 (bases 1 to 8189)  
 AUTHORS Lelong,C., Glaser,P., Presecan,E. and Danchin,A.  
 TITLE Bacillus subtilis gerB downstream sequence  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 8189)

```

/protein_id="CAB07470.1"
/db_xref="GI:1894768"
/db_xref="SPTREMBL:P66739"
/translation="MKFVKAIWFEVAIVFEMSAFKNDOLITDOEKOKIDWEMNKI
OOQEEPVNANK"
2880..2914
/gene="ywtC"
3000..4258
/gene="ywtD"
3000..3007
/gene="ywtD"
3017..4258
/gene="ywtD"
/notes="product similar to E. coli NLPC protein and to
Listeria species p60-related proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB07476.1"
/db_xref="GI:1894769"
/db_xref="SPTREMBL:P66740"
/translation="MNTLANWKKFLVAVIICFLVPIIMTKAFIAEADTSSELIVSEAK
NLGQYKYGETPKGEFDPGLIYVFSKADIHLPKRVNDQYKIGTAVKPKENLKPGD
ILFFKKEGSTVPTPHDALYIGDQMVHSTGSKGVIITNYKKSSVMSCGYICARIAA
DPATADVPVQAEKYIGVPYVFGGSTPEGEFDCSGLVQYVFOALGIYILPSAEQQW
AVEKVAPONIKPGDVVFSNTYKFGISHAGIYAGAGFIQASERKVTIISYLSDEYD
KSKMTGIRPFNDLTPKENPIVSEATLIVGEVPYKQGGVTPTGTAGTGVQYVYQKA
AGISLPATSOYNAGTKIERADLKPGLDGVFFPOSTSLNPSIYIGNQVHVHTLSNGVT
ITNMNTSTYMDKYAGSIRVQ"
4259..4293
/complement(4292..5170)
/gene="ywtE"
/complement(4292..5152)
/gene="ywtE"
/notes="product similar to Bacillus subtilis yxeH and ycsE
proteins and to E. coli yida protein"
/codon_start=1
/transl_table=11
/protein_id="CAB07471.1"
/db_xref="GI:1894770"
/db_xref="SPTREMBL:P66741"
/translation="MKCIAIDLTGLNKEISVAENREAIKRAVDAGILVITICTGRA
TFDKALLDDIDIPITAAANGDTIHDTGVRLSRTLMDQEAQKATADYLSKNIPYEVY
TDDHLLSPDCEAKLHAEFLDLTKSANPNEOTDDLMQAGTQFGKIPKPIHESVED
GGYNYKLICFSFDMOKLQAKELKHKHLAQTSKKHITFELIPASSGKGRALTKLA
DIYGLFQDIIAIGDSFNDLSWFEVAGHRIANENAIDELKEKSTFVTKSNDENGVAIF
IDQLLSGOYA"
/complement(5163..5170)
/gene="ywtE"
5277..5305
/gene="ywtF"
5277..5305
/gene="ywtF"
/complement(5308..6291)
/gene="ywtF"
/complement(5308..6276)
/gene="ywtF"
/notes="an in frame TGA stop codon is found at position
5599 to 5601"
/translation similar to Bacillus subtilis LytR protein"
/codon_start=1
/pseudo
/transl_table=11
/db_xref="PID:e1301651"
/complement(6285..6291)
/gene="ywtF"
6594..8112
/gene="ywtG"
6594..6601
/gene="ywtG"
6609..7982
/gene="ywtG"
/notes="product highly similar to metabolite transport
proteins"

```

```

/codon_start=1
/transl_table=11
/protein_id="CAB07473.1"
/db_xref="GI:1894771"
/db_xref="SPTREMBL:P66742"
/translation="MKKOSNIWLYFFGALGALNYGYDTGVISGAILFMKKELGLNAPT
BGLVSSLLVGNILCSGAAGKLDRFGRKKALMAALLFCIGGLGVALAPNTGVNVLF
RIILGLAVGTSTIIVFLYSELAPKRRKGLSSNLQMLTGYILLSYIVNTYIFADAEA
WRMIGLAAVPSLLLLIGILFMPESPRLFTNGEESAKKILKLRGKTDQIDQIHDH
KEAEKQDEGKLEKLFDPWVRPALIAGLAFLOQFIGTNTIYYVAPKFTTNNVGFNSA
SILGTVGIGTIVLMTLVAKIIDKIGRKPILLFENAGMVISLIVLALVNLFPNTPA
ASWTTVICLGVIIVVEAVSMGVPVWMLPELPLHVRGIGTGVSTLMHVTLLVSLT
YPIIMEAIGISYLIYAIGIMAFIFVRPKVYETKGRSLSEIEQDLKRNKGQGAAG
KQQRVGR"
7981..8112
/gene="ywtG"
/complement(8019..8189)
/gene="gerBC"
/complement(8019..8189)
/gene="gerBC"
/codon_start=1
/transl_table=11
/product="gerBC"
/protein_id="CAB07474.1"
/db_xref="GI:1894772"
/db_xref="SPTREMBL:P66743"
/translation="LQHEYKIDPVFLSDHRIQHPDYWNKVKGHWDEIFSETDPKYDI
SFKIINFGTVGR"
2410 a 1673 c 1857 g 2249 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 626.50      Length: 531
  Ratio: 1.976        Gaps: 8
  Percent Similarity: 59.699   Percent Identity: 30.132

alignment_block:
US-09-652-292-2 x BSZ92954 ..

Align seg 1/1 to: BSZ92954 from: 1 to: 8189

17  LeuGlyGlyLeuThrPheGlyTyrGluLeuAlaValIleSerGlyAlaLe 33
6651  CTTGAGGCGCGCTTATATGGCTATCATCCGAGTGATTCGCGAGCTAT 6700
33  uLeuProLeuGlnLeuAspPheGlyLeuSerCysLeuGluGlnGluPheL 50
6701  TTTATTTATGAAAAAGAGGAGTAGGCTTAACGCGCTTTACAGAAAGGCTTG 6750
50  euValGlySerLeuLeuLeuGlyAlaLeuAlaSerLeuValGlyGly 66
6751  TTGTCAGCTCCTTCTGCTGGTGGCGCATATTTGGGCTCAGGAGCGCGCGC 6800
67  PheLeuIleAspCysTyrGlyArgGlyGlnAlaIleLeuGlySerAsnLe 83
6801  AAGCTGACTACCGCTTTCGGAAGAAAAAACAATATATGGCAGCGCGCT 6850
83  uValLeuAlaGlySerLeuThrLeuGlyLeuAlaGlySerLeuAla 100
6851  GCTGTTTTGTTATAGCGCGCTTGTGTGGTGGTGGTGGTGGTGGTGGTGGT 6900
100  rpLeuValLeuGlyArgAlaValValGlyPheAlaIleSerLeuSerSer 116
6901  TCATGGTGCTGTTTCGGCATCATTTTGGGACTTGCGTGGGAAACATCGAGC 6950
117  MetaLacCysCysIleTyrValSerGluLeuValGlyProArgGlnArgG 133
6951  ACAATCGTACCCCTTATTTATCTGAATCGGCGCAACATAACACGCG 7000
133  yValLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeu 150
7001  GCGCTGTATCATCACTGAATCATGATGATGATGATGATGATGATGATGATG 7050

```

432	rPLeuValLeuSerGluIleTyProValGluIleArgGlyAAlaPhe	448			
7684	GGGTGATCGTCTCTCGTAATTTGTTCCCGCTTCACGTCAGAGGAATCGCGACC	7733			
449	AlaPheCysAspSerPheAsnTTPaAlaAlaAsnLeuPheIleSerLeuSe	465			
7734	GGTGTTTCGACCTTAATGTTTCGACGTTGGGACACTGATGTTTCATTAAAC	7783			
465	rPheLeuAspLeuIleGlyThrIleGlyLeuSerTTPhrPheLeuLeuT	482			
7784	CTATCAATATTAAATGAAAGCGATCGAATCAGTATTTATTATTCTCTGATTT	7833			
482	yrGlyLeuThrAlaValLeuGlyLeuGlyPheIleTyLeuPheValPro	498			
7834	ATGCGCGCGATCGGTATCATGCGGCTCTTATTTGTCGATTTAAAGTGACA	7883			
499	GluThrLysGlyGlnSerLeuAlaGluIleAspGlnGlnPheGlnLysAr	515			
7884	GAGACAAGGGAAGACCTTGAAGAAATTGACGAGGATTTACGGGACAA	7933			
515	gargPheThrLeuSerPheGlyHisArgGlnAsnSerThrGly	529			
7934	AAACGCGACGGGAGGCGCGCGGAAACAGCAGACGTGTCGA	7976			
seq_name: gb_ba:BSUB0019					
seq_documentation_block:					
LOCUS	BSUB0019	212610 bp	DNA	BCT	24-JUN-1999
DEFINITION	Bacillus subtilis complete genome (section 19 of 21): from 3597091 to 3809700.				
ACCESSION	Z99122 AL009126				
VERSION	Z99122.1 GI:2636029				
KEYWORDS	.				
SOURCE	Bacillus subtilis.				
ORGANISM	Bacillus subtilis				
REFERENCE	Bacteria: Firmicutes; Bacillus/Clostridium group;				
AUTHORS	Bacillus/Staphylococcus group; Bacillus.				
	1 (bases 1 to 212610)				
	Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brana, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Chel, S.K., Codan, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emmerson, P.T., Entian, K.J.D., Errington, J., Fabrot, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golighly, E.J., Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Haech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holtsappel, S., Hosono, S., Hullo, M.F., Ilaya, M., Jones, J., Joris, B., Karamata, D., Kashner, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinols, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Owada, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portelle, D., Porwollik, S., Prescott, A.M., Prescan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serrero, P., Shin, B.S., Solido, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tanakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambut, R., Wedler, E., Wedler, H., Weltzenger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.				
The complete genome sequence of the gram-positive bacterium					
Bacillus subtilis					
Nature 390 (6657), 249-256 (1997)					
JOURNAL MEDLINE 98044033					
REFERENCE 2 (bases 1 to 212610)					

# AUTHORS TITLE JOURNAL

Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

Direct Submission  
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,  
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724  
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,  
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45  
68 89 48

## FEATURES

Location/Qualifiers  
1. .212610  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
complement(10. .3879)  
/gene="yvnB"  
complement(10. .3879)  
/gene="yvnB"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15521.1"  
/db\_xref="GI:2636030"  
/db\_xref="SPTREMBL:O34986"  
/translation="MRKVTVIASILLFLSVLSGHHESKAPVYVQELFQTPHYPL  
LENPPQIDVKAEMRSNLMLEAQVRDYSQAQSAFLFYKQSNELGKMWPMPTPGAVR  
KFLAOPKRLIWNSELEYIIVMSNGKQVSETRKLLKLEGYQADLAHPELLITELAV  
TFVWGRADGTEFLFNTDRTIDFYHIRYRYPKEGPDSDLIWRDPERILPSGE  
DLSVAPKRLSSVLYRPLNGTAELNISIGENKPSGIVLOAQVPDOKRKIK  
LDKEPVIEDITDRKVPRAESIELRADIRDSLVKTVAFYRTDENKPKRIAEKD  
RNDLPHYIVYSPELIGKQLEYVVAAGDGNEARTPVKMDIDIKOTSKAHLRWNIE  
RDTLSGTOFLKATTEGRADSIKLDWKKQVTEPAKEKEVTFADFRTKNIYFKNAV  
MEGKVLKVFDDTNNKYRYSVLPETLLRKQKQLOIRITRSQSVSPDFAENRDDL  
VKNARLVDSGTIVRQVRSPEKELFGDNQORSNKQSFQNLPGDGLTSQLEWDTLS  
KLSEGAHHIQSDKSNVLRVNDNSGPHLEPNITEQYKGNLIQADMYDKWSRI  
EEAASLDESITLPIYHTSSDLDLPGHSLKVTATDLAGNKTVIERFKTEREPDOP  
EVIDSEADTHAKLSVRYKQPTDAMDVGFRGQYARNHVLEKHASLTERPKSTV  
PERETPTNKELERSAAGKTVSTENKELFYRFEVTVDPDSIDNDLAEVWKGSS  
LPGRKVTYANNYENWQPVSDVADKDFETLKASVIAADFVRESKMWVVODEIP  
PAKMDYTFWMSDQTYAESYPHFADKQTEWIKQNKQALNKYVFTGDIIVDSADIR  
QNRKADMSVLDKSGIPYGLAGNHGDKGYSYRAFGYFGSDREDKFKHYGGSYK  
NNRGHYDLISSNGDITMYNGAGTIDEDTAMNQVLKHPDRMAILAFHEYLVSNG  
SPIGEKIFKEIVRPNPNVNVLSGHVSAHRKTDLEDGDCGKPDRLVHQLADYGG  
GPEGGQVRLRLQFQDQANDMVHSTYSPYVKDKNYDPTGYGNKDEFSLDLKPRIK  
KVETDYFCNVTYNEELGKREOUKSGDTAEFRDHLPEQSVYTWIIVDSFNKTKS  
PIWKFTKKETRYRPAQDPQDFRHSNP"  
4044. .4517  
/gene="yvnA"  
4044. .4517  
/gene="yvnA"  
/function="unknown"  
/note="similar to hypothetical proteins from B. subtilis"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15522.1"  
/db\_xref="GI:2636031"  
/db\_xref="SPTREMBL:O34692"  
/translation="MDNSIHDELFOALQOFPALKRDKRWKQVQIPSIGISSQHHDHLK  
KMDTLQHLIVSCIHHSQNVNNSFLASRLHISAASVKAHVHLEKHNITVTYKPKGNK  
KEIFYTLDSGRKLAALHEQLEHAKAEQYKQLPNEFSIDDLKTVTAFFNLWIKYM"  
complement(4558. .5775)  
/gene="cypX"  
complement(4558. .5775)  
/gene="cypX"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450-like enzyme"  
/protein\_id="CAB15523.1"  
/db\_xref="GI:2636032"  
/db\_xref="SPTREMBL:O34926"  
/translation="MSQSKLFSVLSQDFONPNPYYAFSOLREEDPVHYEESIDSVFIS  
RYHDVRYILOHPDIFTKTSLVERAEPVPMRGVLAQMHGKHSARRIYRVSIGDALD  
HLSPLIKQNAENLAPYLRGKSLYNDGKTFACVCTMDGLGDKRDHEKISEWHSQ  
VADFITSLSQSEARAHSLWCSEQLSQYLLMPVIKRRVNPNGSDLSILCTSEYEGMAL

SDKDLTALLINVLNLAATSPADKTLALMTYIHLNPNPEQNDVLADRSVPRAIAETLRY  
KPPVQLPRQSQDITVVGGMETIKDITVFCMGAANDRPEAFQPDVFNIIHREDIGIK  
SAFSAARHLAFAGSGIHNCVGAFAAKNEIEIVANIVLDKMRNIRLEDFCYAESGLYT  
RGPVSLLVAFDGA"  
complement(5791. .6537)  
/gene="yvmC"  
complement(5791. .6537)  
/gene="yvmC"  
/function="unknown"  
/note="similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15524.1"  
/db\_xref="GI:2636033"  
/db\_xref="SPTREMBL:O34351"  
/translation="MTGMVTRRSVHFAEALTENCHEIFERRRHVLVIGSPFNSRFS  
EDYIYRLIGNAKQFQSVSVLAGHEAANLLEALGTPRGAERKVRKEVSRNRRFAER  
ALVAHGDPKRAIHTFSFDIDNKAYOLLQROVEHAFEPQPHRACLDMKREAIIGRAR  
GSLMMEEVSEDLNLAVEVYIAELPFFIGAPDILEVEETLLAYHRPWKLGEKLSNHE  
FSCIRPNQGLIIVQEAQMLSEKRITSEG"  
6963. .7472  
/gene="yvmB"  
6963. .7472  
/gene="yvmB"  
/function="unknown"  
/note="alternate gene name: yzhA  
similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15525.1"  
/db\_xref="GI:2636034"  
/db\_xref="SWISS-PROT:P40762"  
/translation="MSDLTKMIYDIYRLLHNEQKANTLSQOFFKEAABEDVAEIP  
KNYTSIHVIDICIGQHPINNAGIARKMNLKANVTIKLKIKEEFINSYQLFDNKKK  
VYFKLTKRGKRGIFFDLHEKLHKKKELAFYQFLDSFQEQKAVLKFLQLTSTLEASQT  
DGTDPDKPVK"  
7493. .8726  
/gene="yvmA"  
7493. .8704  
/gene="yvmA"  
/function="unknown"  
/note="similar to transporter"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15526.1"  
/db\_xref="GI:2636035"  
/db\_xref="SPTREMBL:O34307"  
/translation="MAHTKSKAVLILYTVCFSAFFASLSQNIYSPILPIKSFHVST  
AMVNLVSVMIVTALMQIILGAIIDFKGARIVLTGYLATAASIGCAVTTDFTLFL  
IFRMIOAGSAALPLIAATTIGOLFTEGNSGAMGTOMLLSVAIAIPVLGGFTGGA  
AGVEGIFWILAAISIVLVNTSITFEKDSPTESMOOAKGNVFAHYSKSIPTNKTGVIL  
TILSFVLFYFAVIVYLPILLTEHYHIDVGIAGLILPLALSTIAGLFLFKRIQAKIG  
LHTLFGSNVIAACSTILPAVTHSVLSVLMTALALFGISMGVIPPLYSTMTITNFEH  
NKGSAIGMFNFIRYTGMAAGPMVSAYLLTMPSAMSPSLGLGFAALSFCLLPPMFSP  
QARTKQKHHM"  
8709. .8726  
/gene="yvmA"  
complement(8732. .9091)  
/gene="yviD"  
complement(8732. .9091)  
/gene="yviD"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15527.1"  
/db\_xref="GI:2636036"  
/db\_xref="SPTREMBL:O34648"  
/translation="MKVWASVILNALLIIVIDGYIDSIHISGIAAIIASLIISILN  
VFLKPLLIIFTLPTMVTGLFLFVINAITLMTASINGDSFQIDGFGTAIWSVILS  
VFHLIIQGIQGLEPLKKK"  
complement(9093. .9290)  
/gene="yviC"  
complement(9093. .9290)

terminator

gene

CDS

gene

CDS

3

```

seq_name: gb_ba:BSUB0021
seq_documentation_block:
LOCUS      BSUB0021      215534 bp      DNA              BCT              26-NOV-1997
DEFINITION Bacillus subtilis complete genome (section 21 of 21): from 3999281
            to 4214814.
ACCESSION  Z99124 AL009126
VERSION    Z99124.1 GI:2636442
SOURCE     Bacillus subtilis.
ORGANISM   Bacillus subtilis
REFERENCE  1 (bases 1 to 215534)
AUTHORS   Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G.,
            Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S.,
            Borriess, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C.,
            Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V.,
            Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J.,
            Daniel, R. A., Denizot, F., Devigne, K. M., Dusterhoft, A., Ehrlich, S. D.,
            Emerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E.,
            Foulger, D., Frit, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
            Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Golligly, E. J.,
            Grandi, G., Guisepi, G., Guy, B. J., Haga, K., Halech, J., Harwood, C. R.,
            Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F.,
            Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
            Klier-Bianchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
            Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Liu, H.,
            Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A.,
            Masuda, S., Maestri, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
            Mizuno, M., Ogawa, K., Ogata, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M.,
            Portetle, D., Porwollik, S., Prescott, A. M., Prescann, E., Pujic, P.,
            Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
            Rivolta, C., Roche, B., Rose, M., Sadate, Y., Sato, T.,
            Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
            Sekowska, A., Seter, S. J., Serror, P., Shin, B. S., Soldo, B.,
            Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
            Takeuchi, M., Tanakashi, A., Tanaka, T., Terstra, P., Tognoni, A.,
            Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
            Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T.,
            Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
            Yoshida, K., Yoshikawa, H. F., Zumbstein, E., Yoshikawa, H. and
            Danchin, A.
            The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
            Nature 390 (6657), 249-256 (1997)
            98044033
2 (bases 1 to 215534)
AUTHORS   Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
DIRECT SUBMISSION
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr.
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
FEATURES             Location/Qualifiers
     1..215534
        /organism="Bacillus subtilis"
        /strain="168"
        /db_xref="taxon:1423"
        complement(11..820)
        /gene="yxiJ"
        complement(11..37)
        /gene="yxiJ"
        complement(47..820)
        /gene="yxiJ"
        /function="unknown"
        /note="similar to gluconate 5-dehydrogenase"
        /codon_start=1
        /transl_table=11
        /protein_id="CAB15933.1"
gene
terminator
CDS

```





```

/translation = "MSGISGLVGMGSNIALNMANKGENVAVNYTRDLTDQLQIK
LDQSLSPYEELEDFVQSLEKPRKIFLMVTAGPVDSVQSLKPLLEEGDVIMDGNS
HYEDTERRYDELKEKGIGYLGVGISGGEVGALTPSIMPGGDRDVYEKAAPILTKIAA

```



133 yValLeuValSerLeuTyrGluAlaGlyIleThrValGlyIleLeuLeu 150  
 30681 GACACTCGGCAAAATGAACAACCTTAATGATTGTCAACGGGATTTTGCTCG 30632  
 150 eTyrAlaLeuAsnTyrAlaLeuAlaGlyThrProTyr...GlyTrpArg 165  
 30631 CCTATATCGTTAACTACCTGTTC.....ACACCGTTGAAGCGTGGCGC 30588  
 166 HisMetPheGlyTrpAlaProAlaValLeuGlnSerLeuSerLe 182  
 30587 TGGATGGTCGGTCTGCTCCACTGCTCGCTCGCTCTCTATTAAATCGGAT 30538  
 182 uLeuPheLeuPro.....AlaGlyThrAspG 191  
 30537 CGCCTTTCATCGCGGAATCGCCAAGATGGCTCGTGAAGCGCGCAGCGAAG 30488  
 191 luThrAla.....ThrHis..... 195  
 30487 AAGAAGCGAGCGCATCATGAACATCATCATGACCCCAAGACATCGAA 30438  
 196 LysAspLeuLeuProLeuGlnGlyGluAlaProLysLeuGlyProG 212  
 30437 ATGGAAATTGGCAGAGATGAAGCAAGGGGAAGCGGAGAAA..... 30399  
 212 YArgProArgTyrSerPheLeuAspLeuPheArgAlaArgAspAsnMeta 229  
 30398 .....AAGAAGACGACACTTGGCGTCTGAAAGCCCAAG...TGGATTTC 30359  
 229 rgGlyArgThrThrValGlyLeuGlyLeuValLeuPheGlnGlnLeuThr 245  
 30358 GCCCTATGCTTCTGATTGGCGTGGCCCTTGCATTTTCCACGACGCGACT 30309  
 246 GlyGlnProAsnValLeuCystTyrAlaSerThrIlePheSerValG 262  
 30308 GGCATTACACGGTCATCTACTACGCGCCCAACCATTTTACGAAAGCGCG 30259  
 262 yPheHisGlyGlySerSerAlaValLeuAlaSerValGlyLeuGlyAla 279  
 30258 TCCTT...GGCACAACGACTTCGGCGCTCGGCACAAATGGGAATCGGCATCC 30212  
 279 aLysValAlaAlaThrLeuThrAlaMetGlyLeuValAspArgAlaGly 295  
 30211 TCAAGCTCATCATGTGTATCTACCTGCCATGATCTTTAATGTACCGTGTCCG 30162  
 296 ArgArgAlaLeuLeuAlaGlyCysAlaLeuMetAlaLeuSerValSe 312  
 30161 CGCAAAAACCTTCATCTCGGGAAGTGTGGGCATCATCACTGACGCTTGC 30112  
 312 rGlyIleGlyLeuValSerPheAlaValPrometAspSerGlyProSerC 329  
 30111 TGCTTTA..... 30105  
 329 ysLeuAlaValProAsnAlaThrGlyGlnThrGlyLeuProGlyAspSer 345  
 30104 .....TCA 30102  
 346 GlyLeuLeuGlnAspSerSerLeuProIleProArgThrAsnGluAs 362  
 30101 GCGGTGCTCTGCACACTGGGGCTTCC..... 30075  
 362 pGlnArgLuproIleLeuSerThrAlaLysLysThrLysProHisProA 379  
 30074 .....GCATCAACAGCG..... 30063  
 379 rgSerGlyAspProSerAlaProProArgLeuAlaLeuSerSerAlaLeu 395  
 30063 ..... 30063  
 396 ProGlyProProLeuProAlaArgGlyHisAlaLeuLeuArgTrpThrAl 412  
 30062 .....TGGATGAC..... 30055

```

412 aLeuLeuCysLeuMetValPheValSerAlaPheSerPhedGlyPheGlyLP 429
      ::::: ||| |::::: :::: :::: :|||
30054 AGTAGTCTTTTCGGTGCTCATATCGTCTTTACAGGCACATGGGGTC 30005
      ||| |::::: :||| :||| :||| :|||
429 rovalThrTrpLeuValLeuSerClutIleYrProValGIluIleArgGly 445
      ||| |::::: :||| :||| :||| :|||
30004 CGGTAGTGTGGGTGCTCATTGCCCTAGCTCTTCCCGTCCAAGACGCGGA 29955
      ||| :||| :||| :||| :||| :|||
446 ArgAlaPheAlaPhecysAsnSerPheAsnTrpAlaAlaAsnLeuPheIl 462
      ||| :||| :||| :||| :||| :|||
29954 GCCGCGACAGGGTTCAGCACTTTGGTGTATTCCGCGGCTAACCTGATCGT 29905
      ||| :||| :||| :||| :||| :|||
462 eSerLeuSerPheLeuaspleulleGlyThrlleGlyLeuSerrThrp 479
      ||| |::::: :||| :||| :||| :||| :|||
29904 GTCGCTGTCTCCCCGTGACTGATGTAAGCGCGATGGCATCGCATGGTCT 29855
      ||| :||| :||| :||| :||| :|||
479 heLeuLeuTyrgLyLeuthAlavalLeuglyLeuGlyPheIIcTYrLeu 495
      ||| :||| :||| :||| :||| :|||
29854 TTATGGTTTTCTCAGTCATTGTGCTCTCTCATTTTTCTTCGGGTCTAC 29805
      ||| :||| :||| :||| :||| :|||
496 PheValPrOgluThrrlysGlyGlInSerLeualagLuileaspGlnlNph 512
      ::::: ||| |::::: :||| :||| :||| :|||
29804 ATGGTACC GGAAACAACCAAGGAAAAGCTTGAGGAAATTGAAGAAGCCT 29755
      ||| :||| :||| :||| :||| :|||
512 egLInLysArg 515
      :::::|||
29754 GAAAAAACGG 29745

seq_name: gb_ba:AF045552

seq_documentation_block:
LOCUS AF045552 5957 bp DNA BCT 12-DEC-1998
DEFINITION Lactobacillus brevis xyl operon and xylose isomerase (xylA),
          xylulokinase (xylB), and D-xylose proton-symporter (xylT) genes,
          complete cds.
ACCESSION AF045552
VERSION AF045552.1 GI:2895853
KEYWORDS .
SOURCE Lactobacillus brevis.
ORGANISM Lactobacillus brevis.
          Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
          Lactobacillus.
REFERENCE 1 (bases 1 to 5957)
          Chaillou,S., Bor,Y.-C., Batt,C.A., Postma,P.W. and Pouwels,P.H.
            Molecular cloning and functional expression in Lactobacillus
            plantarum 80 of xylT, encoding the D-xylose-H+ symporter of
            Lactobacillus brevis.
JOURNAL Appl. Environ. Microbiol. 64 (12), 4720-4728 (1998)
MEDLINE 93054905
REFERENCE 2 (bases 1 to 5957)
          Bor,Y.-C. and Batt,C.A.
            The D-xylose operon of Lactobacillus brevis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5957)
          Bor,Y.-C. and Batt,C.A.
            Direct Submission
TITLE Submitted (30-JAN-1998) Food Science, Cornell University, 312
        Stocking Hall, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source           1..5957
                     /organism="Lactobacillus brevis"
                     /db_xref="taxon:1580"
     gene             1230..5860
                     /gene="xyl operon"
     misc_feature     1230..1253
                     /gene="xvl operon"
                     /note="xvl operator"
     gene             1271..2688
                     /gene="xyla"
     CDS              1271..2620
                     /gene="xyla"
                     /codon_start=1
                     /transl_table=11
```

<p> <code>/product="xylose isomerase"</code>  <code>/protein_id="AAC95125.1"</code>  <code>/db_xref="GI:2895854"</code>  <code>/translation="MT E E Y K W G D K I Q Y V G H Q D K K S G L G F Y Y N P E E T M G K K M K D L C P R F A V A Y W H T F D O R L V D P G D G T A G R P Y D K Y T D P M D L A L A K V D A A F E Y O K L G V D Y L C F H O R D L A P E G D T L E R T N A L D K V D K I V E Y O K T S G M K V L H N T S M F T N P R P E V G A A T S P Y A D V P A Y S A Q L K H S L E I G K R V G S E N Y F W G G R E G Y S L E S L W N T N M K Q E Q H A A K I F H M A K D Y A N E I G D A M L L E P K P E P T T H Q Y D F A A T T A F M K E Y D L D K D F K L N L E G N H A N I A G H T Y Q H E I T R A R E A K P L A S L A N O G D K L I D E Y P S N L Y E T T A A M Y E V Y V E N G S S P R G L G D P A K P R A F P E D L F L I V G M D F S F A A G L R V A A A M K Q D C F L D N L K A D R Y S S Y K G V G A D I E S C K A D L K S L E A V A I D K P O S E L I A A T H S D H L E E I K T I N H Y I I D T L S K Y 2661..2688</code> </p>		terminator
<p> <code>/gene="xylA"</code>  <code>/note="potential Rho-independent transcription terminator"</code>  2724..4287  <code>/gene="xylB"</code>  2724..4232  <code>/gene="xylB"</code>  /codon_start=1  /transl_table=11  <code>/product="xylulokinase"</code>  <code>/protein_id="AAC95126.1"</code>  <code>/db_xref="GI:2895855"</code>  <code>/translation="MCKYVLVGLGTSAVKVSALDSHGQIVAQESFDYDLIOKOPCYN E Q N P D W S C T T V A I V R L I L N D H L A S N I E G L S Y S C M G H L V L D E N K V L R P A L I L N D T S R Q R E I E A K L A D E F V H I T R N Q P L E G F I T L K L W Y K Q N E P D I W A K R T F V L P R K Y R Y R T M G N L A M D Y S A T G T V L L D V A K E M S Q R I C A A L D P M S M C P P L I K S I D L A G V T P A Y A P F S G I T T D T F G G A A N A G A G A G I L H P N M V L S S I G T S G V L K Y E D N A D V N Y G V A L O P E H A I P D K F Y S G M V T A A G Y S F T W F K K T P A P A E D T D V V A S A A K S T V A N G L L V T P Y I G E R A P A D A I R G S F T G V D G T H Q R V D F V R A L E G I I I S F R D L F D I Y E N G G D F T V S I G G G A K S P L M L Q I A D I F N R K V S L T N E Q G P G M G A A M L A A T G L G F D S L Q D C A E T F V H R G K A Y E P N D N V K Y E K M H A I Y K V Y Q O T K T I S E Q L L D Y R R A E L " 4248..4287</code> </p>		terminator
<p> <code>/gene="xylB"</code>  4344..4367  <code>/note="potential Rho-independent transcription terminator"</code>  <code>/gene="xyl operon"</code>  <code>/note="xyl operator"</code>  4454..5880  <code>/gene="xylT"</code>  4454..5827  <code>/gene="xylT"</code>  /codon_start=1  /transl_table=11  <code>/product="D-xylose proton-symporter"</code>  <code>/protein_id="AAC95127.1"</code>  <code>/db_xref="GI:2895856"</code>  <code>/translation="MRKVSSTGYFFFGALGGLFGYDTGVTGIVSGAILFIQKMNLAGSQW Q V H V A G A A L I P T Y L A E L A P S D K R C T V S L P O L V M Y T G I L L A Y I T N Y S F S C F Y T R I I L G M A V G A S A L I P T Y L A E L A P S D K R C T V S L P O L V M Y T G I L L A Y I T N Y S F S C F Y T G W R M L G P A R I P A N L I L F G L L I P E S P R E L Y R G S H L D E A R H V L D P N K H Q D Q A V A K E I N D I Q E S A K I V S G W S E L F G M W R P S L I I G I G L A I F O O V M G C N T V L Y I A P T I F T D V G F G S Q T A A I S V I A L T V Y I A F F S A T W G P V M V M I G E V F P L N I R G L N S F A S V I N N T A N N I V S L T F P S L L D F F G T G S L F I G V G I L C F A S I W F V Q K K V P E T R N R S L E D I E A T L A K T G S D A A E L S T T K " 5842..5880</code> </p>		terminator
<p> <code>/gene="xylT"</code>  <code>/note="potential Rho-independent transcription terminator"</code>  BASE COUNT 1600 a 1244 c 1444 g 1669 t  ORIGIN </p>		
<p> alignment_scores:  Quality: 571.00 Length: 515  Ratio: 1.936 Gaps: 7  Percent Similarity: 57.282 Percent Identity: 27.379 </p>		
<p> alignment_block:  US-09-652-292-2 x AF045552 </p>		
<p> Align seg 1/1 to: AF045552 from: 1 to: 5957 </p>		



186 roAlaGly..... 188  
 1009 CTGAAGCCCTCGATGGCTTATTACAGAAAGACAGACTCAGAGGCCGT 1058  
 189 .....ThrAspGluThrAl 193  
 1059 AGNAATTTATCTCAGATGGCTGTACACAGACCATTTGATGAGGAATATGA 1108  
 193 aThrHisLysAspLeuIleProLeuGlnGlyGluAlaProLysLeuG 210  
 1109 TAGCATCAAAACAACATTGAAGAGGAGGAAAGAGGTTGGCTCAGCTG 1158  
 210 lProGlyArgProArg...TyrSerPheLeuAspLeuPheArgAlaArg 225  
 1159 GACCTGTGATCTGCAGAAATGCTGAGTTATCCCCCAACTGCCCGAGCT... 1205  
 226 AspAsnMetArgGlyArgThrValGlyLeuGlyLeuValLeuPheG 242  
 1206 .....TTAATGTGGGTTGTGGCTTACAAATGTTCCA 1237  
 242 nGlnLeuThrGlyGlnProAsnValLeuCysTyrAlaSerThrIlePheS 259  
 1238 GCAGCTCTCAGGCATTTAACACCATCATGTACTACAGTGAACCATTTCTGC 1287  
 259 erSerValGlyPheHisGlyGlySerSerAlaVal...LeuAlaSerVal 274  
 1288 AGATGCTGGTGTGAAGATGATGAGACTTGCAATATGGCTGGCTTCAGTT 1337  
 275 GlyLeuGlyAlaValLysValAlaLaThrLeuThrAlaMetGlyLeuVa 291  
 1338 ...ACAGCCTTACAAATTTTTCACACTTGTGGGAGTCTGGCTTGT 1384  
 291 lAspArgAlaGlyArgArgAlaLeuLeu.....LeuAlaGlyCysA 305  
 1385 TGAGAAGGTGGCCGCAGAAAGCTTACCTTGTGTAGTTTAGCAGGT...A 1431  
 305 lLeuMetAlaLeuSerValSerGlyIleGlyLeuValSerPheAla... 320  
 1432 CCACCGTAGCACTTATTTCTTGGCTTGGGATTTGCTATCAGGCCAA 1481  
 321 .....ValProMetAspSerGlyPr 327  
 1482 GTTTCCTCCAGGCATCACTTTTAAGCAATAGCTCCGTGAGTCRGAACGC 1531  
 327 oSerCysLeuAlaValProAsnAlaThrGlyGlnThrGlyLeuProGlyA 344  
 1532 CACTTCACAAAGATAC...AGTTACTGTAATGAATGTATCTTGGATCCAG 1578  
 344 spSerGlyLeu.....LeuGlnAspSerSer 352  
 1579 ACTGCGGTTTCTGCTACAAGATGAACAAATCAACTGTCTATGACTCTCTCC 1628  
 353 LeuProProIleProArg.....ThrAsnGluAsp..... 362  
 1629 TGTGTTCCAGTTAATAAGCATCTACAAATGAGGCAGCCCTGGGGCAGGTC 1678  
 363 .GlnArgGluProIleLeuSerThr.....AlaLysLys 374  
 1679 TGAATAATGAACCAAGCTCAAAACACAGAGATATATTTTGGGCTTACAAT 1728  
 374 hLysProHisProArgSerGlyAspProSerAlaProProArgLeuAla 390  
 1729 TCTGCCCTACTCCATCTCC..... 1748  
 391 LeuSerSerAlaLeuProGlyProProLeuProAlaArgGlyHisAlaLe 407  
 1748 ..... 1748  
 407 uLeuArgTTPThrAlaLeuLeuCysLeuMetValPheValSerAlaPheS 424  
 1749 .....TGGACTGCACCTCTGGGCCCTTATTTTATATCTGTCTCTTTC 1791

424 exPheGlyPheGlyProValThrTrpLeuValLeuSerGluIleTyrPro 440  
 1792 CACCTGGAATGGGACCAATGCTTGGACTGTGAAATTTCTGAAATATATATCCC 1841  
 441 ValGluIleArgGlyArgAlaPheAlaPheCysAsnSerPheAsnTrpAl 457  
 1842 CTTTGGGCAAGAAGTACAGGAATGCATGTTTCATCTGGAATAAATGGAT 1891  
 457 aAlaAsnLeuPheIleSerLeuSerPheLeuAspLeuIleGlyThrIleG 474  
 1892 TTTCAATGCTGCTGTTTCACTAACATTTTACACACAGCAGAGTATCTTA 1941  
 474 lLeuSerTrpThrPheLeuLeuTyrGlyLeuThrAlaValLeuGlyLeu 490  
 1942 CATACTATGGAGCTTTCTCTCTATGCTGATTTGCTGCTGGGACTC 1991  
 491 GlyPheIleTyrLeuPheValProGluThrLysGlyGlnSerLeuAlaG 507  
 1992 CTTTTCATCTATGCTGCTCTCTGAGACCAAGGCAAAAAATTAGAGGA 2041  
 507 uIleAspGlnGlnPheGlnLysArgPheThrLeuSerPheGlyHisA 524  
 2042 AATTGAATCACTCTTTGACAAACAGGCTATGTACATGTGGCACTTCAGATT 2091  
 524 rgGlnAsnSerThrGlyIleProTyrSerArgIleGlu 536  
 2092 CTGATGAAGGGAGATATATTGGAATATATATCGCGGTGAAG 2129



**THIS PAGE BLANK (USPTO)**